Three Essays in Development Economics

by

Jessica E. Leight

B.A., Yale University (2006)
M. Phil., Oxford University (2008)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Author ....
Department of Economics
May 15, 2013

Certified by ....
Abhijit Banerjee
Ford International Professor of Economics
Thesis Supervisor

Certified by ....
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Thesis Supervisor

Accepted by ....
Michael Greenstone
Chairman, Department Committee on Graduate Studies

Three Essays in Development Economics

by

Jessica E. Leight

Submitted to the Department of Economics
on May 15, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis seeks to analyze two questions central to the economic welfare of rural households in developing countries: the trade-offs between equity and efficiency induced by hybrid forms of property rights, and the decisions made by households in the process of human capital accumulation, particularly in early childhood. The first two chapters examine the impact of periodic land reallocations and tenancy reforms in China and India, respectively, on intravillage land inequality and the investment decisions made by rural households. The final chapter turns the focus to inequality within the household, analyzing how intrahousehold allocation of educational resources in rural China responds to inequities in endowment between children.

In the first chapter, I evaluate the impact of village-level land reallocations in China on household economic outcomes. Since land was decollectivized in China in 1983, village leaders have implemented regular forced reallocations of land designed to enhance intravillage equity and attain other policy goals. I estimate the impact of insecure tenure using the past history of land shifts as an instrument for current tenure insecurity, and find that an increase in the probability of losing the current plot yields a decrease in agricultural inputs and production of around one standard deviation. Though the costs of insecure tenure are high, structural estimates of the varying cost of reallocation across different villages suggest the choice to reallocate does reflect an optimizing process on the part of village officials, who reallocate where the net benefit is larger. However, the observed pattern of reallocations would be optimal only given an objective function for the village leader that places an extremely high weight on equity, and even given this objective function, there is evidence that village leaders may be making some costly mistakes.

In the second chapter, coauthored with Timothy Besley, Rohini Pande and Vijayendra Rao, we seek to analyze the long-run impact of land reform in southern India. Though land reform policies have been widely enacted across the developing world, evidence about the long-run impact of these policies remains quite limited. In this paper, we provide evidence about these long-run effects by combining the quasi-random assignment of linguistically similar areas to South Indian states that subsequently pursued different tenancy regulation

policies with cross-caste variation in landownership. Roughly thirty years after the bulk of land reform occurred, land inequality is lower in more regulated areas, but the impact differs by caste group. Tenancy reforms increase own-cultivation among middle caste households, but render low caste households more likely to work as daily agricultural laborers. At the same time, an increase in agricultural wages is observed. These results are consistent with credit markets playing a central role in determining the long-run impact of land reform: tenancy regulations increased land sales to the relatively richer and more productive middle caste tenants but reduced land access for poorer low caste tenants.

In the final chapter, I analyze the strategies employed by households in rural China to allocate educational expenditure to children of different initial endowments, examining whether parents use educational funding to reinforce or compensate for these differences. Empirical results obtained employing early-childhood climatic shocks as an instrument for endowment, measured as height-for-age, indicate that parental expenditure is preferentially directed to the relatively weaker child. In response to the mean difference in endowment between siblings, parents redirect between 10 and 20% of discretionary educational spending to the child with lower endowment, and this effect is robust across multiple measures of endowment and multiple measures of climatic shocks. This analysis is consistent with the hypothesis that parents use the intrahousehold allocation of resources to compensate for differences in endowment and thus in expected welfare, between their children.

Thesis Supervisor: Abhijit Banerjee
Title: Ford International Professor of Economics

Thesis Supervisor: Esther Duflo
Title: Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics

Acknowledgments

I would like to begin by acknowledging with utmost thanks the guidance, advice and support of my advisers, Abhijit Banerjee, Esther Duflo and Rohini Pande. Their expertise, patience and seemingly unlimited knowledge were instrumental in propelling me through not only this dissertation, but my entire graduate education, and the breadth and depth of their understanding of development economics never ceases to amaze. I am immensely grateful for all their help.

Warm thanks are also due to Panle Jia for her outstanding assistance. Despite my status as a novice in structural estimation, she patiently guided me through the production function literature and provided valuable support and encouragement as the project progressed. In an earlier phase of the project, Stephen Ryan also assisted greatly in the development of a framework to analyze the objective function of village leaders. I would like to thank him as well for his generous feedback and assistance.

I would also like to thank Timothy Besley, Paul Glewwe, Ben Olken, Albert Park and Tavneet Suri for useful feedback on various chapters of my thesis. Particular thanks are due to John Giles, as well as Albert Park and colleagues at the Gansu Survey of Children and Families for sharing data used in the empirical analysis. Huge thanks to fellow or former graduate students Melissa Dell, Ben Feigenberg, Horacio Larreguy, Laura Ralston, Jennifer Peck, and Xiao Yu Wang, for many hours (and years) of discussion, feedback, assistance and entertainment.

I am grateful to the NBER Health and Aging Fellowship for funding.

Last but not least, I would like to acknowledge many, many years of love and support from my entire family: my parents, parents-in-law, and brothers (particularly my brother Willie, fellow traveler through many years of education), and most of all my husband Amir. It wouldn't have been possible without you.

Contents

Chapter 1

Reallocating Wealth? Insecure Property Rights and Agricultural Investment in Rural China

1.1 Introduction

The establishment of clear land rights has long been considered a key milestone in the development of the modern industrialized countries. Because land is the principal asset in a preindustrial economy, the development of an institutional structure that encourages its efficient use is argued to enhance growth substantially (North & Thomas 1973). Conversely, the absence of stable and enforced property rights is widely identified as a major impediment to growth in today's developing countries (De Soto 2000).

Despite this emphasis on the importance of private property rights, however, collectively owned or managed land remains a widespread phenomenon in the developing world. Collective or partly collective land structures continue to be predominant in rural areas in China, in Mexico, and in many parts of sub-Saharan Africa. These forms of land ownership can yield substantial benefits in terms of equity, but they may also generate significant efficiency costs.

In China, the post-Mao period saw the emergence of a hybrid system of landownership, in which formal title to land is held by the village collective and use rights are held by households. Moreover, plots are subject to periodic reallocations between households conducted by village officials every three to five years, thereby generating systematic insecurity in land tenure. These reallocations represent the outcome of a bargaining process between officials and households that weighs the costs and benefits of the associated disruption in property rights. The objective of this paper is to estimate the economic costs of insecure land tenure induced by these periodic reallocations, and to demonstrate that village officials do respond to variation in these costs in shaping the relative security of local property rights.

First, I examine variation in tenure insecurity within a village conditional on a reallocation being conducted. Households that have recently had their land reallocated are less likely to have their land reallocated in a subsequent round, and accordingly the past history of changes in landownership can be employed as an instrument for the probability of loss of the current plot. The results show that the reduction in the probability of losing the current plot as a result of past inclusion in a reallocation, a decrease of around 5% on a base probability of 56%, results in an increase in the use of agricultural inputs and in total agricultural output of between .05 and .1 standard deviations, with no evidence of simultaneous substitution out of non-agricultural activities.

This effect of relatively more secure tenure is evident for households that both gained and lost land in the past. Accordingly, any plausible alternative channel for the observed pattern would require that an unobservable shock correlated with reallocation affect both relatively rich and relatively poor households in the same way relative to the mean. The observed pattern of symmetric increased investments by households at both ends of the landownership distribution in the year of a reallocation is inconsistent with most obvious sources of omitted variable bias.

Second, while these reduced form estimates capture a uniform effect of reallocation, the observed variation in the propensity to reallocate across space and time suggests that the costs and benefits of reallocation are in fact far from uniform. It is plausible that officials will choose to reallocate in areas where disruption of land tenure is less costly — more specifically, less costly in terms of investment foregone as a result of insecure land rights.

To test this hypothesis, I estimate an agricultural production function that allows for spatial variation in the returns to agricultural inputs. I find that the propensity to reallocate is negatively correlated with returns to lagged inputs, and thus with the magnitude of the investments lost as a result of the induced insecurity of tenure. Despite the fact that reallocations generate large costs, the observed pattern of reallocations does seem to reflect an optimizing process on the part of the village official.

In the final part of the analysis, I postulate a functional form for the objective function that underlies this optimizing process and seek to estimate parameters for this function that would best reproduce the observed pattern of reallocations. The results suggest first, that reallocations are only optimal for a village leader that places an extremely high weight on increased equity relative to the potential output losses induced by a reallocation. Second, even given an objective function that values equity, village leaders are making some potentially costly errors by reallocating where the relative balance of benefits and costs is unfavorable.

To sum up, the evidence from variation in property rights in China suggests that even incremental shifts in the security of land tenure in a context of partly collective land rights can have large economic implications. In addition, variation in the frequency of reallocations is correlated with their costs, with reallocations occurring most frequently where they are least costly. Thus at a local level, property rights adapt to reflect the relative returns to secure property rights in different economic environments.

This paper supplements an existing literature that has evaluated the impact of varying regimes of property rights in China. Feder, Lau, Lin & Luo (1992) argue based on a before-and-after analysis that excessive investment in nonproductive assets such as housing is evidence of the negative impact of insecure land tenure. Brandt, Huang, Li & Rozelle (2002) analyze the impact of land tenure by comparing households' private plots, assigned permanently to households in some villages for their personal cultivation, with responsibility land that is subject to reallocations. Similarly, de la Rupelle, Quheng, Shi & Vendryes (2009) use household-level heterogeneity in land rights within a village to identify the impact of reallocations on outmigration, finding that insecure land rights induce temporary, rather than permanent, outmigration in order to ensure claims are retained on land left behind. Both papers make the assumption that plots are exogenously assigned to different contractual types within a village.

Jacoby, Li & Rozelle (2002) analyze the impact of insecure tenure on investment in rural China by using a hazard model to estimate predicted risks of expropriation for different plots held. They find that a higher expropriation risk decreases investment in organic fertilizer, employing the identifying assumption that the hazard of reallocation is exogenous to household characteristics.

There is also a larger literature about the economic impact of property rights that evaluates land reforms in which tenants without formal title are endowed with stronger property rights (Banerjee, Gertler & Ghatak 2002, Besley & Burgess 2000). Goldstein & Udry (2008) analyze property rights in Ghana and conclude that individuals with more secure tenure rights by virtue of their more powerful political positions invest more in maintaining soil fertility. Another set of papers focused on urban land policy in Latin America finds that land titling increases labor supply and investment (Besley 1995, Field 2005, Galiani & Schargrodsky 2010). In the historical literature, Hornbeck (2010) analyzes the impact of the introduction of barbed wire on agricultural productivity in the western U.S., and concludes that the stronger protections of land title afforded by barbed wire led to a significant increase in settlement, land values and crop productivity.

This paper adds to this existing literature while making a number of new contributions. First, I evaluate the impact of insecure tenure on an unusually large set of economic outcomes. Second, I demonstrate a systematic correlation between the frequency of local disruptions to property rights and variation in the costs of those disruptions, estimated using an agricultural production function. Third, I estimate parameters of the objective function corresponding to the village leader's choice to reallocate. Given that China has been the site of some of the most far-reaching experiments in property rights over the last fifty years, evidence about both the political economy and the economic consequences of insecure property rights in rural China can be a useful contribution to the ongoing debate about how to structure land rights to maximize rural growth (Deininger & Binswanger 1999).

The remainder of this paper proceeds as follows. Section 1.2 provides an overview of the relevant institutional background. Sections 1.3 and 1.4 provide a brief conceptual framework for the analysis and describe the data. Section 1.5 presents the results focusing on intravillage heterogeneity in security of tenure, while Section 1.6 analyzes cross-sectional variation in the costs of reallocation derived from an estimation of the agricultural production function. Section 1.7 discusses estimation of the village leader's objective function, and Section 1.8 concludes.

1.2 Background

Property rights in China have a long and tumultuous history in the post-1949 era, and the institutional framework that governs rural households remains unusually complex. This section provides a broad overview of the history of property rights during the Communist period, as well as the characteristics of the periodic reallocations that have been a feature of the rural land ownership system since 1983.

1.2.1 Property rights under the Household Responsibility System

Since 1983, land rights in China have been characterized by a system of collective land tenure in which partial use rights are assigned to the household, a system widely known as the household responsibility system. Prior to this, agricultural production in China between 1962 and 1978 was largely organized around the institution of the production team, a unit consisting of 20 to 30 households that jointly farmed agricultural land and sold the resulting output, distributing the associated income to participating laborers according to a system of workpoints intended to reward labor, skill and political commitment.[1] The overarching imperative of agricultural policy during this Maoist period was to maximize grain production in order to feed urban areas and support industrialization drives, a goal enforced using substantial mandatory production quotas and low procurement prices. By 1978, the cumulative impact of these policies was disastrous, leading to low rural income, land degradation and a severe undersupply of non-grain crops (Walker 1984).

As a result, the new government led by Deng Xiaoping, acceding to power shortly after Mao Zedong's death, introduced major changes in agricultural policy. First, the household was reinstated as the primary unit of agricultural production under a system variously known as household contracting or the household responsibility system. Each household was provided with an allocation of land for its own use, while land title continued to be held by the village. The household also committed to the delivery of a fixed amount of quota grain sold to the state at a preset price, in addition to taxes owed.

Excess production could then either be sold to the state at a higher, above-quota price, or at rural markets (Lin 1992), with the household having full rights over residual, post-quota income. Households were also allowed control over a private plot of land used to cultivate crops other than grain or to raise animals; income from this plot accrued entirely to the household (Walker 1984). The average per capita land endowment was small, less than one fifteenth of a hectare, and a household's endowment generally comprised several fragmented parcels (Wen 1989).

At the same time, major adjustments were made to the state's system of agricultural targets and agricultural procurement. Prices for government procurement of most agricultural goods, previously so low that they often did not cover costs, were raised substantially. In addition, a previously elaborate system of targets for sown area, inputs, production and yield for a variety of agricultural productions was simplified to government procurement targets for key agricultural goods only (Lin 1992).

These changes were implemented in a piecemeal fashion between 1979 and 1983, beginning with a few isolated provincial or local experiments, and subsequently spreading widely to a point of almost total decollectivization by the end of 1983 (Unger 1985). The establishment of the household responsibility system led to a substantial increase in the growth rate

[1] The team farming system was itself a retreat from the much larger agricultural communes formed during the Great Leap Forward between 1958 and 1962, in which land and labor were collectivized in communes of 6,000-8,000 households (Chinn 1979).

of agricultural output, which had been only 2% annually over the previous 25 years. Between 1978 and 1984, agricultural output increased nearly 8% annually. One analysis estimated that roughly half of this growth was due to increased use of inputs, particularly fertilizer, and half to the assignment of land use rights to households (Lin 1992).

1.2.2 Land reallocations

However, property rights under the household responsibility system remained crucially incomplete, principally because land was subject to periodic land reallocations. The stated aim of these reallocations was to promote equity in land ownership, and to adjust landholdings in response to changes in household size.[2] However, the policy clearly created an opportunity ripe for rent-seeking by local officials (either local government officials or Party leaders, known as cadres).

Accordingly, the literature has observed that "it is not uncommon that a few village cadres or officials choose to conduct readjustments simply in order to exert their influence and authority for other dubious purposes" (Keliang, Prosterman, Jianping, Ping, Reidinger & Yiwen 2007). Another analysis noted that the threat of reallocation was frequently used as a carrot and stick to ensure compliance with other administrative goals (family planning targets, grain quotas, corvee labor obligations, and taxes) relevant to local leaders' opportunities for promotion. Leaders employed the threat of land reallocation to induce households to comply with other policy goals and minimize their enforcement costs, or to punish households for an absence of compliance (Rozelle & Li 1998).

At the same time, reallocations required considerable investment of time on the part of village leaders, entailing "countless discussions and negotiations among village cadres and the involved households pertaining to the new land assignment exercise" (Kung 2000, Brandt et al. 2002). To cite a specific example, a survey in July-August of 1999 found that a third of villages that had decided to carry out a reallocation in accordance with a land law passed the previous August had still not implemented it (Schwarzwalder, Prosterman, Jianping, Riedinger & Ping 2002). Though reallocations normally occurred at the end of the year during the fallow winter period, the lapse in time required for implementation introduced scope for strategic behavior, for example hastening the marriage of sons (or delaying the marriage of daughters) in order to maximize the number of family members in the household

[2]Given that variation in the number of children is limited by virtue of the One Child Policy, the relevant changes in household size are normally driven by marriage of adult children. Daughters typically exit the household, while daughters-in-law will arrive. Other changes might be driven by migration, death, or changes in extended family structure.

when its required allotment of land was determined (Unger 2005).

A larger literature in both economics and political science has assembled descriptive evidence about the frequency and nature of land reallocations over time. Brandt et al. (2002) find that there is a negative correlation between the frequency of reallocations and the number of plots per household, as well as the total number of households in the village. Kung (2000) uses a separate survey of land reallocations and notes that reallocations decline in frequency when terrain is more rugged or hilly, and when landholdings are more scattered or fragmented. Unger (2005) also documents the negative relationship between topography and reallocations and finds a negative relationship between the frequency of reallocations and the availability of off-farm income-earning opportunities.

The central government has made periodic attempts to regulate reallocations. By the 1990s, national policymakers became increasingly concerned that insecure tenure was the primary reason for a decline in agricultural growth rates relative to the early years of the Household Responsibility System. As a result, a (nonbinding) policy directive was issued in 1993 establishing a fixed term of land tenure equal to thirty years. This policy was then embodied in law in 1998, requiring that land be contracted to households for 30 years. Readjustments during this period were still allowed, but needed to be approved by two thirds of village members; villages were also allowed to conduct a reallocation immediately after the introduction of the new policy.[3] The law also mandated the issuance of written contracts or certificates to farmers.

Despite the seeming boldness of this reform, subsequent survey evidence indicated that its implementation was extremely mixed. A majority of farmers continued to express low confidence in their tenure security and believed subsequent reallocations were inevitable (Schwarzwalder et al. 2002). A later law in 2002 outlawed reallocations completely except in extreme cases and spelled out the right to lease, exchange and carry out other land transactions, excluding sale and mortgage. This reform is, however, beyond the chronological scope of this analysis (Keliang et al. 2007), which will focus on the impact of reallocations on rural economic outcomes between 1987 and 2002.[4]

[3] A survey in 1999 reported in Schwarzwalder et al. (2002) inquired about whether villages had decided to conduct such a reallocation and whether it had taken place, the source of the previously cited data about the delay inherent in the implementation of reallocations.

[4] Data from the household survey employed here is not publicly available for the years after 2002; accordingly, it is not possible to employ data from the post-2002 period in a placebo test. There is, in addition, an ongoing debate about how well these subsequent reforms were implemented and thus how secure property rights in the post-2002 period are.

1.3 Conceptual framework

1.3.1 Optimizing reallocation

Consider the decision made by a village leader of whether or not to undertake a reallocation in a given village in a given year. A reallocation has both costs and benefits. The advantages may include private benefits for the official in rent extraction or future opportunities for promotion, as well as quasi-public benefits such as an increase in equity that may also be valued by village households.

On the other hand, reallocations also have costs. Households that are uncertain about their long-term tenure on a given plot will not make investments whose returns accrue partly in the medium-term, thus resulting in a decline in agricultural investment and output; a simple model of a household production function demonstrating this result is presented in Appendix 1.A. These costs are clearly highly salient to households. For simplicity, I will assume here that officials are themselves indifferent to this loss in output. They are, however, forced to take into account the preferences of households by bargaining over whether or not to hold a reallocation.

Assume that the official and each household face a variant of the single-seller, single-buyer problem; they need to bargain over the sale of a single good, a reallocation of land. The official places a value B on this reallocation, capturing benefits that include opportunities for rent-seeking and decreased intravillage inequality.

Each household i in the village places a value on a reallocation that can be written as follows, equal to the negative of the value of continued land tenure $\bar{v}$ plus the value of the expected change in land $w(E[\Delta L])$. For simplicity, I assume that every household in the village would have its land tenure disrupted by the reallocation.

$$v_i = -\bar{v}_i + w(E[\Delta L_i]) \tag{1.1}$$

$\bar{v}_i$ is defined more specifically as the loss in output due to foregone investments that are not made when tenure insecurity is introduced by a reallocation. Note that X_i^{NR} denotes a vector of agricultural investments made by household i in the absence of a reallocation, while X_i^R denotes investment in the case of a reallocation.

$$\bar{v}_i \equiv F(X_i^{NR}) - F(X_i^R) \tag{1.2}$$

Some households may place a negative value on reallocation if they face significant losses due to reduced long-term investment, and thus they will seek to avoid a reallocation. Others may place a positive value on reallocation if they expect to gain land in the process. Each household has the option to impose a bargaining or lobbying cost c_i on the official in the case of the outcome they do not prefer: i.e., a household that prefers a reallocation be avoided can inflict a lobbying cost at the time of the reallocation, and vice versa for a household that prefers a reallocation.[5]

Total bargaining costs are summed across all households in the case of a reallocation, defined $C(R=1) = \sum_i c_i(R=1)$, or a non-reallocation, defined $C(R=0) = \sum_i c_i(R=0)$. There is also a transactional cost of time and effort T needed to redefine land boundaries. This transactional cost is assumed to be higher for localities with more rugged topography; this assumption is consistent with the prior literature, as well as the intuition that implementing a land swap perceived to be fair is more challenging in areas with variable topography and thus more local heterogeneity in land quality.

The village official will reallocate if the benefits exceed the sum of bargaining and transaction costs:

$$B > C(R=1) - C(R=0) + T \tag{1.3}$$

Accordingly, the variable R_{vt}, defined as equal to one if a reallocation occurs in village v in year t and zero otherwise, can be viewed as a function of benefits of the reallocation for the official, its costs in lost output, and the topographic characteristics of the village.

$$R_{vt} = f(B_{vt}, C_{vt}, T_v) \tag{1.4}$$

This conceptual framework suggests that villages where households place a larger value on continued land tenure, i.e. $\bar{v}_i^h = F(X_i^{NR}) - F(X_i^R)$ is greater, should also exhibit a lower frequency of reallocations. In these villages, the cost in terms of foregone output of tenure insecurity is greater, and accordingly households will bargain more aggressively against reallocations.[6]

Intuitively, in some villages there may be few profitable long-term investments available.

[5]This framework assumes that households can commit to imposing a certain cost on village officials. While this is clearly a strong assumption, it could be easily nested in a multi-period model in which households that fail to impose the postulated bargaining penalty on the official suffer a loss of credibility in future bargaining rounds.

[6]In addition, none of the preceding analysis precludes the possibility that the official himself also faces a direct loss from foregone output, via lowered tax revenue or other channels. In this case, the direct benefit B of reallocation will be lower in villages where the lost output as a result of tenure insecurity is larger; this serves only to strengthen the postulated negative correlation between the output costs of reallocation and their frequency.

A reallocation will decrease the probability that households in the village will make such investments because they face the risk of losing their plot prior to the next growing season, and thus losing any lagged returns to this year's investments. However, this may not be a significant loss if these lagged returns are low in magnitude. More specifically, the difference in investment and thus in output between the reallocation and the non-reallocation case is increasing in the returns to lagged agricultural inputs, a comparative static also demonstrated in Appendix 1.A. If the returns to lagged inputs are higher, the loss in investment as a result of a reallocation is higher, the net benefit of reallocation for the official is lower, and accordingly reallocations should be observed less frequently.

1.3.2 Optimizing household-level land shifts

To sum up, the observed distribution of reallocations across villages and years can be understood as the outcome of a complex bargaining process that leads to some officials choosing to conduct reallocations in certain years while others do not. However, village leaders who have chosen to hold a reallocation then face another set of optimization decisions: how and to whom to reallocate land within the village. Some households will gain or lose land, while other households may not see changes to their landholdings.

The probability that a given household i in village v and year t will see its land reallocated is denoted D_{ivt}; it is assumed to be a function of household characteristics X_{ivt}, conditional on $R_{vt} = 1$. If there is no reallocation, then $D_{ivt} = 0$ for all households.

$$D_{ivt} = \begin{cases} f(X_{ivt}) & \text{if } R_{vt} = 1 \\ 0 & \text{if } R_{vt} = 0 \end{cases} \tag{1.5}$$

Potential household covariates X_{ivt} relevant to the reallocation decision could include demographic characteristics that render the household a poor match with its current land allotment; the household's current position in the overall distribution of landownership, given the village leader's interest in equity; and the past history of land shifts for the household.

Accordingly, there are two sources of variation in insecure tenure for the households of interest, corresponding to two separate optimization margins for the village official. There is variation in the household probability of land shifts D_{ivt} conditional on a reallocation occurring ($R_{vt} = 1$), corresponding to the official's choice of which households to reallocate. There is also variation in the probability of reallocation across villages and years, corresponding the official's choice of whether or not to hold a reallocation. In this analysis, I will exploit both sources of variation in tenure insecurity.

1.4 Data

The dataset employed here is a panel collected by the China Research Center for the Rural Economy (RCRE), comprising a sample of 299 villages in 13 provinces in China every year between 1986 and 2002, excluding 1992 and 1994. Figure 1-1 shows the sample counties. A randomly selected sample of households in each surveyed village forms the panel; the mean number of households in a village-year cell is 69. Summary statistics are shown in Table 2.7.

Measures of land reallocation are constructed using household reports of changes in their household landholdings from year to year, excluding land leased.[7] A shift in landholdings is identified at the household level if a household reports a change in land area owned of at least .1 mu, where a mu is the Chinese unit of land area (comprising .165 acres).[8] A reallocation is defined to have occurred when the proportion of households reporting a change in their landholdings in a given village in a given year exceeds the 75th percentile across all village-years or the proportion of land reported transferred exceeds the 75th percentile across all village-years. This definition is employed to exclude those cases where a small number of households report a change in landholdings as a result of measurement error or a private contractual arrangement that is not sanctioned by the village leadership.

Figure 1-2 shows histograms for both measures used to define reallocations. Both show a spike close to zero and a long right tail with a higher proportion of transfers. The reallocation measure employed captures this right tail. In addition, the first stage and the reduced form are robust to altering this definition, and results employing varying definitions of reallocations will be shown in the robustness checks.[9]

Past literature on reallocations that has estimated their frequency has largely used data drawn from two sources: surveys of village leaders, e.g. Kung (2000), or surveys of individual households conducted periodically by the Rural Development Institute that obtain retrospective statistics over a long recall period (Schwarzwalder et al. 2002). Survey data

[7] There is no uniform policy regarding the legality of land leasing arrangements in rural China. In this sample, leasing is rare; only around 8% of household-year observations report any land leased in or out. Leased land is thus of limited relevance to the rural economy overall.

[8] .1 mu represents around 2% of median land owned.

[9] This definition makes no distinction between different types of plots that households may hold (e.g., responsibility land versus private plots); though the dataset reports limited information on holdings of responsibility land, inputs and agricultural production are not reported by type of plot. Accordingly, the resulting estimates should be viewed as mean effects across all household landholdings.

of leaders has the advantage of employing a clear definition of reallocation. However, leaders may also face incentives to bias reports of reallocations toward zero to avoid reporting reallocations that are not in line with national land policy guidelines. Retrospective data collected at the household level, on the other hand, may be imprecise and biased by recent events.

While survey data of leaders indicate that reallocations occur around every 5 years (Kung 2000), the reallocation measure constructed here R_{vt} shows reallocations occurring around every three years. It is plausible that a measure based on household reports of land shifts will be noisier and thus more likely to generate spurious reports of land reallocations, a source of classical measurement error. However, this strategy for identifying household reallocations has the additional advantage of allowing the direct examination of the changes in landholdings at the household level that were induced as a result of the reallocation. A measure constructed from reports by village leaders, by contrast, provides no information about the mechanics of the implementation of the reallocation within the village.

1.5 Intravillage heterogeneity in security of tenure

1.5.1 First stage

In analyzing reallocations, I will begin by considering variation in tenure insecurity within a village, taking as given the observed distribution of reallocations across different villages and years. When a reallocation does occur in the sample villages, ex ante all households face the risk of the suspension of their use rights and the transfer of their plot. However, not all households experience a change in landholdings in every reallocation.

In order to evaluate the effect of variation in security of tenure on economic outcomes, it is useful to begin by analyzing the characteristics of households that do have their land reallocated. Assuming that the quantity of land already held is of first-order relevance, I first estimate the probability of a household's land being reallocated conditional on a reallocation occurring in the village for households in each decile of landownership. These probabilities are shown graphically in Figure 1-3.

The evidence indicates that land transfers are broadly progressive. The probability of receiving a positive transfer of land via a reallocation is generally decreasing by decile, and the probability of a negative transfer is increasing. Only the tenth (and richest) decile appears to be somewhat insulated from the effects of reallocations. Otherwise, households from the lower deciles are generally more likely to gain land, and households from the upper deciles more likely to lose it.

Now, assume one reallocation has already occurred in every village in the past. Both reallocation "winners" and reallocation "losers" have experienced a shock to their landholdings and, presumably, to other economic outcomes as well. Two groups of households can be defined based on whether their land was affected in the last reallocation: $DP_{ivt}^{-1} = 1$ defined for household i in village v in year t denotes a household that gained land in the previous reallocation (on average, three to five years prior), and $DN_{ivt}^{-1} = 1$ denotes a household that lost land. These households have received opposite shocks, relative to the unaffected households, with the median (absolute) change in landholdings observed as a result of a reallocation around one third of median land owned.

There is, however, one characteristic common to all households that had their land reallocated in the previous round: a decline in the probability that their land tenure will be disrupted again in the next reallocation. Reallocating land for a household incurs fixed transaction costs. Accordingly, it is logical to assume that village leaders will seek to minimize the number of land transfers they effect over time, conditional on reaching their goals of equity or an improved match between households and land, and this in turn implies that a series of incremental land transfers is unlikely to be welfare-maximizing for the official. Instead, they would seek to fully adjust a household's land to its optimal level when a reallocation is implemented, implying a lengthy period until either subsequent demographic shocks render the household's landholding suboptimal, or the household is again due for an equity-enhancing shift in land.

To test this hypothesis, I estimate the impact of past reallocation inclusion on a dummy variable capturing inclusion in the current reallocation, denoted D_{ivt} for household i in village v in year t. D_{ivt} is defined to be equal to one if a household reports any change in total land owned above the threshold (.1 mu) in the year of the reallocation. R_{vt} is defined as equal to one if a reallocation is observed in village v in year t. The independent variable of interest is a dummy variable for a household's past reallocation inclusion; I estimate the effect of this variable on D_{ivt} in years in which reallocation is observed. A control for each strata of landownership prior to reallocation L_{ivt} and village and year fixed effects are included.[10]

$$D_{ivt} = \beta_1 DP_{ivt}^{-1} + \beta_2 DN_{ivt}^{-1} + \beta_3 L_{ivt} + \nu_v + \gamma_t + \epsilon_{ivt} \tag{1.6}$$

This equation is estimated for the household panel post-1995, to allow for coding of D_{ivt}^{-1}

[10] L_{ivt} is an integer variable controlling for each vingtile (5%) of landownership, i.e. ranging from 1 to 20. This controls for the position of the household in the overall distribution of landownership.

based on prior reallocations reported in the first section of the panel. 1995 is chosen as the cut-off year as a new and more comprehensive survey of household economic outcomes was administered for the first time in that year. The coefficients on β_1 and β_2 from estimating (1.6) are shown in Column (1) of Table 3.4, and are negative and significant. In other words, inclusion in a reallocation leads to a significant decrease in the probability that a given household will have its land adjusted again in the next reallocation for both past reallocation gainers and past reallocation losers, compared to households that were not reallocated. This is consistent with the intuition that multiple, sequential reallocations of land for the same household are unlikely to be optimal.

These results suggest that past reallocation history generates a plausible source of quasi-exogenous variation in current tenure insecurity. The implied exclusion restriction is that a reallocation has no differential impact on households that were included in a past reallocation and households that were not included, other than via the channel of differential probability of current reallocation and thus differential tenure insecurity. Column (2) shows the result of estimating the same equation with past reallocation participation pooled across gainer and loser households. The pooled dummy for past inclusion in a reallocation is denoted D_{ivt}^{-1}.

$$D_{ivt} = \beta_1 D_{ivt}^{-1} + \beta_2 L_{ivt} + \nu_v + \gamma_t + \epsilon_{ivt} \tag{1.7}$$

This is the first stage relationship of interest, and the same negative and significant relationship is evident.

In the two-stage least squares analysis, I employ the same specification using the full sample of both reallocation and non-reallocation years, and add a full set of interactions with R_{vt}. The coefficient on D_{ivt}^{-1} when $R_{vt} = 0$ is zero by construction, and thus the primary coefficients of interest in the first stage are identical. Having had land reallocated in the past is correlated with the probability of a disruption to current land tenure, but only if a reallocation is actually occurring in the village; otherwise, the impact of past reallocation on current reallocation is precisely zero.

Moreover, the heterogeneity of past reallocation patterns (including both gainers and losers) can be used as an additional test of the exclusion restriction. Any bias in unobservables as a result of past reallocation-induced shocks to land is presumed to be of opposite sign for past land gainers, who now own more land than the mean household, and land losers, who now own less land. Figure 1-4 shows estimated kernel densities of landownership for households with past positive and negative shocks to landholdings in reallocations, partialling out village and year fixed effects. Both a shift right in the distribution for past land gainers and a shift left for past land losers are evident.

Accordingly, if the reduced form impact of past reallocation status on economic outcomes in a reallocation year is observed for both past land gainers and past land losers, this suggests that the observed effect is plausibly interpreted as a causal estimate of the impact of tenure security on economic outcomes. A violation of the exclusion restriction would require that reallocation is correlated with a shock that affects both the relatively land-poor and the relatively land-rich, a non-monotonic pattern that would seem a priori implausible. Further evidence about the validity of the exclusion restriction will be presented in the next section.

1.5.2 Reduced form and 2SLS

The reduced form specification is the following, where Y_{ivt} denotes economic outcomes at the household level. Controls for lagged household reallocation D_{ivt}^{-1}, strata of landownership L_{ivt} and the full set of interactions with reallocation R_{vt} are included.

$$Y_{ivt} = \beta_1 D_{ivt}^{-1} \times R_{vt} + \beta_2 D_{ivt}^{-1} + \beta_3 L_{ivt} + \beta_4 L_{ivt} \times R_{vt} + \nu_v + \gamma_t + \nu_v \times R_{vt} + \gamma_t \times R_{vt} + \epsilon_{ivt} \quad (1.8)$$

The reduced form can also be estimated as the "split" reduced form, including both DP_{ivt}^{-1} and DN_{ivt}^{-1} and the corresponding interactions as explanatory variables. The 2SLS specification is the following, where D_{ivt} is a dummy for forced reallocation of land at the household level, instrumented by $D_{ivt}^{-1} \times R_{vt}$.

$$Y_{ivt} = \beta_1 D_{ivt} + \beta_2 D_{ivt}^{-1} + \beta_3 L_{ivt} + \beta_4 L_{ivt} \times R_{vt} + \nu_v + \gamma_t + \nu_v \times R_{vt} + \gamma_t \times R_{vt} + \epsilon_{ivt} \quad (1.9)$$

The assumed timing in each year is as follows: a signal about the reallocation is received prior to household's investment decisions. Investments are made and output is realized. Subsequently, land is reallocated after the harvest.[11] While the exact timing of the reallocation decision vis-a-vis household investment decisions doubtless varies, the assumption is that the considerable time required to implement a reallocation requires a decision to be made at a point that overlaps with the period of key investments, in line with the evidence that households are observed to respond strategically to early notifications about future reallocations in decisions about household formation and marriage. Such strategic behavior would be impossible if the decision to reallocate was simultaneous with the actual implementation.

Eight outcome variables are reported for each specification: land cultivated, fertilizer,

[11] There is considerable anthropological evidence that reallocations normally occur during the fallow period in winter. See for example (Unger 2005).

agricultural labor, a dummy for agricultural structures, moveable capital, grain production, and dummies for labor in a non-agricultural household business and for labor outside the household. Land cultivated, fertilizer, agricultural labor and grain production are normalized by the area of land owned prior to the reallocation; all variables are then normalized by the mean and standard deviation of the outcome variable in the control (non-participating) households.[12]

Reduced form results Panel A of Table 3.5 shows the results from estimating the reduced form, and Panel B the "split" reduced form with dummies for both past reallocation gainers and losers, $DP_{ivt}^{-1} \times R_{vt}$ and $DN_{ivt}^{-1} \times R_{vt}$ respectively. Note again that all dependent variables are normalized to have mean zero and standard deviation one.

In Panel A, the coefficients on the interaction $D_{ivt}^{-1} \times R_{vt}$ are generally positive and significant with magnitude between .05 and .1, reflecting greater agricultural investments by households that were included in the last reallocation and accordingly enjoy greater tenure security. No effect is observed for moveable capital, labor input into household businesses or labor in outside enterprises. This is consistent with the intuition that the returns to moveable capital (an index of animals, tools and machines owned) and non-agricultural activities are unaffected by reallocations.

In the split reduced form, the coefficients are positive and significant for households that gained and lost land in the past. The fact that the estimated coefficients are generally slightly larger for past losers is consistent with the evidence of a larger first stage for these households (i.e., their relative tenure security is greater). However, the final row of Panel B reports a test of equality of the coefficients β_1 and β_2 on $DP_{ivt}^{-1} \times R_{vt}$ and $DN_{ivt}^{-1} \times R_{vt}$, and the hypothesis that the coefficients are equal is uniformly not rejected.

In addition, the assumption that there is no omitted channel that is biasing the estimated effect for both land losers and land winners can be tested by examining the estimated coefficients on DP_{ivt}^{-1} and DN_{ivt}^{-1}. In general, these coefficients are of opposite sign, though not statistically significant; households that gained land seem to employ less inputs per acre and are somewhat less likely to have non-agricultural businesses. Most importantly, the absence of any pattern of symmetric and significant coefficients on the dummies for past reallocation gainers and losers suggests there is no common bias in observables across both

[12]Fertilizer is defined as the mean of total fertilizer and the most common subtype of fertilizer used, carbamide. Moveable capital is defined as the sum of animals and tools; agricultural structures is equal to one if a household reports any agricultural structures or associated machines. The top 5% and bottom 1% of observations of each continuous outcome variable are trimmed to remove the influence of outliers. The asymmetry reflects the much longer right tail in the distribution of agricultural input variables.

sets of households. The only exceptions are positive coefficients on the probability of outside labor.

The exclusion restriction for the instrumental variables analysis requires that there is no shock correlated with a village-level reallocation R_{vt} that differentially affects households included in the previous reallocation. If households that had their land previously reallocated either positively or negatively showed characteristics that were significantly different from households with no previous reallocations in non-reallocation years, this would suggest that the past history of reallocations generated different trends for both land winners and land losers. Furthermore, if there were an interaction between these trends and R_{vt}, the coefficients of interest would be biased. Given the absence of any evidence of significant difference in outcomes for past reallocation participants in non-reallocation years, however, it seems implausible that there is another, independent shock correlated with R_{vt} that affects only these households in reallocation years. Further evidence about differing trends for past land gainers and land losers in the years prior to a reallocation will be presented in the robustness checks.

2SLS results To reiterate the key assumptions underlying the two-stage least squares result, the exclusion restriction requires that a reallocation has no differential effect on households that were and were not included in the previous reallocation, other than via the channel of differential tenure security. Given the asymmetric nature of past reallocation inclusion, encompassing both reallocation winners and reallocation losers, the necessary assumption can be further refined: there is no shock correlated with a reallocation that affects both relatively land-poor and relatively land-rich households compared to the mean.

Under this assumption, Table 1.4 shows the results from estimating equation (3.13), the instrumental variables specification. The coefficients indicate that households facing the mean probability of losing their plot in a reallocation year (around .6) exhibit a decline in area sown, fertilizer, agricultural labor and total agricultural production, all around .8 standard deviations in magnitude. There is no significant change in moveable capital and no change in the probability of non-agricultural employment. While there is weak evidence of a decline in agricultural structures, the estimated effect is not significant.

These results are consistent with a model of household behavior in which households decrease the use of inputs that have medium-term returns and inputs that are complementary to those medium-term investments. The shift in sown area may reflect a decline in the prevalence of multicropping.[13] Optimized multicropping yields long-term benefits in terms of soil nutrition and health (Zhang, Shen, Li & Liu 2004), and thus households expecting short tenure may be less likely to multicrop. The decline in both multicropping and fertilizer use generates a decline in agricultural labor, presumably a complementary input.

On the other hand, no effect is observed for non-agricultural activities. Given that both the establishment of a non-agricultural household business and the search for outside employment (often rationed in rural China) may require considerable initial, and potentially irreversible, investments, it would be implausible to see a substantial divergence in non-agricultural investments between households with different short-term expectations of land tenure. No such divergence is observed. This evidence is consistent with the hypothesis that the observed impacts represent the effect of variation in short-term insecurity in land tenure, rather than other unobserved differences between households with different reallocation histories.

Panel B of the same table shows the results of estimating equation (3.13) controlling for a quadratic polynomial in land area held by each household, also interacted with R_{vt}. This specification tests whether differences in plot size between households that did and did not participate in past reallocations are a source of bias, and the estimated coefficients are consistent and in fact more precise. Panels C through E show the two-stage least squares results where the sample is restricted according to certain criteria. These specifications are discussed below in the robustness checks.

1.5.3 Robustness checks

This section presents a series of robustness checks on the above results.

Differing trends for households with different past reallocation histories It can also be hypothesized that households included in the last reallocation, who enjoy relatively greater tenure security, begin to show higher investments in years prior to the subsequent reallocation. This pattern could emerge for two reasons: first, while the previous specification assumed that households have perfect information about the timing of a reallocation, in fact this information may be noisy. Households may perceive some latent risk of a reallocation occurring in the year or two prior to its actual date. Second, even if they perfectly anticipate

[13]The average household in this dataset is multicropping around 50% of land owned, a rate consistent with previous estimates from agricultural censuses and remote sensing data (Frolking, Qiu, Boles, Xiao, Liu, Zhuang, Li & Qin 2004).

the next reallocation date, they may begin to taper down investments that have a time horizon longer than the anticipated time lapse to the next reallocation.

The objective of this robustness check is to evaluate whether the difference in agricultural investments between households that previously participated in a reallocation (who have relatively greater tenure security) and households that did not (who have less tenure security) is evident in years prior to the implementation of a reallocation. In order to test this hypothesis, the reduced form equation (1.8) is re-estimated for the reallocation year (denoted T=0) and each year leading up to a reallocation. For simplicity, the variable R_{vt} and associated leads R_{vt}^{+1} (one year prior to the reallocation), R_{vt}^{+2} (two years prior to the reallocation), etc. enter the equation linearly rather than interacted with village and year fixed effects.[14] Thus the equation of interest can be written as follows, for example for the first lead R_{vt}^{+1}:

$$Y_{ivt} = \beta_1 D_{ivt}^{-1} \times R_{vt}^{+1} + \beta_2 D_{ivt}^{-1} + \beta_3 L_{ivt} + \beta_4 R_{vt}^{+1} + \nu_v + \gamma_t + \epsilon_{ivt} \tag{1.10}$$

The estimated parameters, capturing the difference in outcomes between households that were included in past reallocations and those that were not in each specified year leading up to a reallocation, are then graphed in Figure 1-5 along with a 90% confidence interval.

The graphs show that for outcomes that are affected by reallocations (fertilizer, sown area, labor, structures and agricultural production), there is generally a pattern of increasing divergence between households previously included in reallocations and households not previously included in the year prior to the next reallocation. Though the estimated coefficients are not statistically significant, they are positive for R_{vt}^{+1}; there is little evidence of a significant trend in longer lags. However, for those outcomes that are hypothesized to be unaffected by tenure security, no significant trend is observed.

These results provide suggestive evidence that households at higher risk of losing their plots may begin tapering their investments in the years prior to a reallocation, though the largest effect is seen in the reallocation year. The absence of any systematic trend in longer lags or for other variables, however, suggests that there are no unobservable characteristics of households previously included in reallocations that are driving the results. In addition, the evidence of a divergence in agricultural inputs between households with different probabilities of future loss of their plot prior to the year of the reallocation suggests that the prior estimates

[14]The reallocation lead variables are coded as follows: moving backwards from the final observed reallocation, each previous year is coded as T=+1, T=+2, etc. When a year with another reallocation is encountered, all the lead variables are re-set to zero. The regression sample size thus shrinks with each additional lead year; for each newly defined sample, the bottom 10% of outliers are trimmed.

of the impact of insecure tenure on investment may be conservative.

Miscoding partial versus full reallocations A second potential challenge to this estimation strategy is the possibility that some of the reallocations identified are what the literature has described as partial reallocations, distinguished by the fact that only households that have had changes in their household composition experience incremental shifts in landholdings, without full swaps of their plots. In this case, households that need more land might receive an incremental, additional transfer, while households that have too much land would lose part of their holdings (Keliang et al. 2007).

In order to address this possibility, there are two separate cases that should be considered. One is that the past reallocation, on the basis of which DP_{ivt}^{-1} and DN_{ivt}^{-1} were defined, is in fact not a full reallocation. The second is that the current event that is generating insecurity in tenure, captured by dummy $R_{vt} = 1$, is not a full reallocation.

Under the first case, some households for which $D_{ivt}^{-1} = 1$ may have previously had their land reallocated partly or primarily because of their changes in composition. If this were the case, then the exclusion restriction required for the instrumental variables specification requires that a reallocation has no differential effect on households that previously experienced a change in composition compared to those that did not, other than via the channel of differing tenure security. A violation of this exclusion restriction would arise if there is a shock correlated with reallocation that differentially affects households with a past history of demographic shifts.

On the other hand, if what we identify as a current reallocation is in fact a partial reallocation or some other type of irregularity in land transfer, and households' expectations are rational, then only some households are subject to decreased tenure insecurity: more specifically, those households that expect to lose land based on a shift in their household's composition. The exclusion restriction implied by this specification requires that there is no shock correlated with land reallocations that differentially affects relatively land-rich (on a per capita basis) households.

In both cases, the exclusion restriction is not the same as that postulated for the primary analysis, primarily because the specification can no longer be interpreted as a symmetric and non-monotonic effect of greater tenure security observed for both relatively land-poor and relatively land-rich households. Accordingly, if miscoding is common and reallocation is correlated with other shocks that affect relatively land-rich households or households with previously unstable composition, this could generate bias. In order to test the robustness of the results to potential bias introduced by the miscoding of partial reallocations, I restrict the sample in several ways.

First, I restrict the sample to households that did not previously report a change in composition in the year of the previous reallocation. These are households that have a history of demographic stability. If the primary results represent bias introduced by correlated shocks for demographically unstable households, this specification should show no significant effect. The results are shown in Panel C, and the estimated coefficients are consistent in both sign and magnitude.

Second, I evaluate the effect of households that can reasonably be assumed not to be relatively land-rich on a per capita basis. If the miscoding of partial reallocations as R_{vt} is common and the estimated effect reflects a correlated shock for these relatively land-rich households, these specifications should show no effect. Panels D and E show the results of re-estimating equation (3.13) restricting the sample first to households that have gained or remained constant in composition (Panel D), and second to households in the bottom half of the land distribution (Panel E). These are households that on the basis of demography and land ownership are plausibly land-gainers, not land-losers. The results again remain consistent in both sign and magnitude, suggesting that there is little bias introduced by miscoding and the potential of correlated shocks for land-rich households.

Alternate measures of reallocation As an additional robustness check, I re-calculate the primary measure of reallocation R_{vt} using alternate definitions based on varying cut-offs in the proportion of households reporting land transfers and the proportion of total land reported transferred. While the primary measure of reallocation employs a cutoff of 75%, I employ a range of cutoffs between 40% and 80% and then use these alternate measures to estimate the reduced form with sown area as the dependent variable, equation (1.8).[15] The reduced form coefficients along with a 90% confidence interval are then graphed in Figure 1-6.

The results show a consistently positive coefficient on the independent variable $D_{ivt}^{-1} \times R_{vt}$ regardless of the cutoff employed, and the estimated coefficients are also significant or close to significant over a wide range of potential definitions of R_{vt}. This suggests that the observed results are not merely an artifact of the definition of reallocation employed.

[15] Note that when R_{vt} is re-defined, D_{ivt} is also re-defined as household-level reallocation inclusion can only vary in a year where $R_{vt} = 1$.

Information as a channel for predicting reallocations The identification strategy also requires the assumption that there are no unobservable characteristics shared by households that experienced both negative and positive shocks to their landholdings in past reallocations that could co-vary with future reallocations. One potentially plausible assumption would be that village officials, who execute reallocations, systematically have more information about households that have similar characteristics to themselves and thus are more likely to participate in their social networks. Given the greater informational salience of these "socially proximate" households, officials may be more likely to alter their landholdings to a level the official regards as optimal. This unobservable proximity to village officials could also generate other time-varying effects if, for example, village leaders prefer to simultaneously implement a reallocation with another policy shift that also differentially affects households with close ties to the village leadership.

This hypothesis can be tested by examining whether there are any characteristics that, when shared with village officials, render proximate households symmetrically more likely to experience positive and negative reallocation shocks to their landholdings. A series of dummy variables are defined that capture households' economic specializations (whether they cultivate rice or wheat, and whether they report household businesses of any of the enumerated types), and a limited number of social characteristics enumerated in the survey (the presence within the household of an individual with education beyond high school, a veteran, resident grandparents or a Communist party member).

For each village-year cell, the mean of this dummy is calculated for households that are reported to be led by a village official, and this official mean is denoted O_{vt}. The equations of interest regress the dummies for positive and negative reallocation in years in which $R_{vt} = 1$ on the household indicator of interest I_{ivt}, the official indicator O_{vt} and the interaction $I_{ivt}O_{vt}$. The equation also includes a control for each strata of landownership L_{ivt} and village and year fixed effects. The specification is thus parallel to the original first stage, where the primary independent variable of interest is the interaction between official and individual characteristics. The objective is to test whether households with a particular economic or social characteristic are more likely to have their land reallocated in villages where officials share this characteristic.

$$D_{ivt}^{P/N} = \beta_1 I_{ivt} O_{vt} + \beta_2 I_{ivt} + \beta_3 O_{vt} + \beta_4 L_{ivt} + \nu_v + \gamma_t + \epsilon_{ivt} \tag{1.11}$$

The results shown in Table 1.5 indicate no clear pattern of coefficients across the various interaction terms. The only household characteristics that seem to generate substantial shifts in both reallocation probabilities are wheat cultivation, husbandry and educational levels, but the effects are not symmetric: when officials engage in husbandry, households that also do so are more likely to gain land in a reallocation and less likely to lose it (i.e., they are favored). Thus it is plausible to conclude that there is very little evidence that informational proximity to village officials serves as a common source of bias for both households that gain land and households that lose land in a reallocation.

1.6 Cross-sectional variation in reallocation costs

If the exclusion restriction for the 2SLS specification estimated above is valid, then the resulting estimates represent the causal effect of a change in the probability of losing the current plot on investment and economic outcomes within the same village and year, conditional on the observed distribution of official reallocation decisions. The estimated cost is uniform for all villages. However, the heterogeneity in the observed probability of reallocation suggests that the benefits and costs of reallocation are far from constant. Moreover, the bargaining process that generates the observed distribution of reallocations is itself a function of these benefits and costs. Accordingly, it is reasonable to hypothesize that there should be a negative correlation between the costs of reallocation and its probability.

The measure for relative costs of reallocation employed here is derived from the model of household optimization outlined in Appendix 1.A. Households are assumed to equate the ratio of returns to labor N_t and fertilizer F_t in agriculture to the ratio of factor prices. In the case of a reallocation, there are no lagged returns to fertilizer and the solution given a Cobb-Douglas production function is standard:

$$F_t = \frac{w_t \alpha_F}{r_t \alpha_N} N_t \tag{1.12}$$

In the counterfactual case of no reallocation, the returns to fertilizer are realized both this period and next period and the ratio of returns to labor and fertilizer has a more complex form. The optimal level of fertilizer solves the following equation equating the ratio of returns to labor and fertilizer (both this period and next period) to the ratio of factor prices.

$$\frac{w_t}{r_t} = \frac{\frac{\partial \pi_t}{\partial N_t}}{\frac{\partial \pi_t}{\partial F_t} + \frac{\partial \pi_{t+1}}{\partial F_t}} \tag{1.13}$$

$$0 = \frac{\frac{\partial \pi_t}{\partial N_t}}{\frac{\partial \pi_t}{\partial F_t} + \frac{\partial \pi_{t+1}}{\partial F_t}} - \frac{w_t}{r_t} \tag{1.14}$$

The first-order condition indicates that when the returns to lagged investment are higher, fertilizer use increases, a result shown algebraically for the Cobb-Douglas case in the appendix. Accordingly, when the returns to lagged investment are higher, the difference in investment between the reallocation and the non-reallocation case and thus the cost of a reallocation is higher — presumably making reallocations less likely.

The objective of this section is to test the hypothesis that the frequency of reallocations is correlated with their relative cost by estimating a production function that allows the returns to agricultural inputs to vary cross-sectionally. First, I will describe the methodology used to estimate the production function. Second, I will present the primary results that test the correlation between returns to lagged investment and reallocations. Third, I will use a difference-in-difference strategy exploiting variation in crop composition over time to examine the robustness of this correlation to the potential endogeneity of agricultural inputs.

1.6.1 Estimating an agricultural production function

The production function postulated is Cobb-Douglas; inputs are labor, land area, fertilizer and lagged fertilizer. Sown area, seeds, labor and output for grain cultivation are reported separately, and fertilizer employed is assumed to be devoted to grain cultivation proportionately relative to its share in total sown area. Lagged inputs are set equal to the amount of that input used in the previous year, provided that the household did not participate in a reallocation last year (i.e., conditional on the household cultivating the same land this year and last year). The objective is to identify lagged returns of inputs on land cultivated continuously by the same household.

First, I estimate the production function using OLS with village, year and crop fixed effects, employing both the full sample and the sample restricted to rice and wheat producers.[16] This specification follows the methodology employed in other production function analyses of Chinese agriculture (Lin 1992, Wan & Cheng 2001). The dependent variable is value added in grain production, equal to grain production valued at the market price minus the cost of seeds.[17] X_{ijt} is the quantity of input j used by household i at time t and $F_{ij,t-1}$ denotes lagged fertilizer; lower-case letters denote log inputs. I focus initially on lagged

[16]In order to identify these households, a dummy for rice or wheat production is set equal to one if a household reports positive sown area for one crop (rice or wheat) and no sown area for the other crop throughout the entire panel.

[17]The standard production function literature has generally employed either value added in production (Olley & Pakes 1996) or revenue (Levinsohn & Petrin 2003) as the dependent variable. The primary specification here employs value added, but in the cross-sectional analysis I will also present results using revenue (the grain harvest valued at the market price) as the dependent variable.

fertilizer because the medium-term returns to this input are most intuitive to estimate and interpret, but I will also estimate the lagged returns to labor, fertilizer and sown area in a subsequent robustness check. μ_c, ν_v and γ_t denote crop, village and year fixed effects.

$$y_{it} = \sum_{j=1}^{J}(\alpha_j x_{ijt} + \alpha_f f_{ij,t-1}) + \mu_c + \nu_v + \gamma_t + \epsilon_{ivt} \tag{1.15}$$

Clearly, ordinary least squares estimates of the returns to agricultural inputs may be biased by the presence of unobserved shocks to productivity. For robustness, the production function is re-estimated using dynamic panel methods as described in Blundell & Bond (2000). A detailed description of the methodology and the results can be found in Appendix 1.B, but in essence, the postulated production function imposes an AR(1) structure on the errors, yielding orthogonality of lagged levels of the independent variables and the error term in the first-differenced equation. Additional restrictions on the correlation between the household fixed effect η_i and differences ΔX_{it} and ΔY_{it} allow for the imposition of additional moment conditions that employ lagged differences as an instrument for the equation in levels. Given that the use of lagged levels as instruments requires dropping observations without observed lags, I estimate the production function only with the full sample of grain producers in order to maintain adequate power. While the primary specification employs the full set of lagged instruments, I also restrict the instruments employed to lags three and four to evaluate the robustness of the results to a change in the instrument set.

The results from the estimation of the production function using both methods can be found in Table 1.6. Each specification is reported with and without the returns to lagged fertilizer. The pattern of coefficients is relatively consistent, with several caveats. The estimated returns to labor are more variable in the dynamic panel regressions and not statistically significant. The point estimates for returns to both fertilizer and lagged fertilizer are larger in the dynamic panel specification, though noisy in the case of lagged fertilizer; however, the difference between the two sets of estimates is not significant. Comparing across all the specifications, the returns to lagged fertilizer are between 10% and 40% of the returns to contemporaneous fertilizer, consistent with the intuition that a non-trivial component of the returns to fertilizer are realized in the medium-term.

The final rows of Table 1.6 report for the dynamic panel specification the results of the Sargan-Hansen test of overidentifying restrictions and the chi-squared test of common factor restrictions imposed in the minimum distance model used to estimate the coefficients. (Again, details can be found in Appendix 1.B.) While the overidentifying restriction is rejected for specifications excluding lagged returns to fertilizer, for the primary specification the validity of lagged levels as instruments is not rejected at the 10% level (Column 3) or at the 5% level (Column 4). The test of common factor restrictions is uniformly not rejected.

1.6.2 Variation in returns to investment and reallocations

In order to test the hypothesis that there is variation in the returns to lagged investment that is correlated with variation in reallocation behavior, I now estimate the production function allowing the returns to inputs to vary by province and crop. Interaction effects between crop fixed effects μ_c and province fixed effects λ_p and all agricultural inputs (x_{ijt} and $f_{ij,t-1}$) are included. The equation is again estimated using both OLS with crop, village and year fixed effects and dynamic panel GMM.

$$\begin{aligned} y_{it} =& \sum_{j=1}^{J}(\alpha_j x_{ijt} + \alpha_f f_{ij,t-1}) + (\sum_{c=1}^{C}\mu_{cj} + \sum_{p=1}^{P}\lambda_{pj}) \times (\sum_{j=1}^{J}\alpha_j x_{ijt}) \\ &+ (\sum_{c=1}^{C}\mu_c + \sum_{p=1}^{P}\lambda_p) \times \alpha_f f_{ij,t-1} + \mu_c + \nu_v + \gamma_t + \epsilon_{ivt} \end{aligned} \tag{1.16}$$

I then calculate the return to lagged fertilizer for each household, corresponding to the linear combination of the returns in the province and for the crop cultivated. This allows for the calculation of the mean return to lagged fertilizer in a given village, which is normalized by the estimated standard error of α_f and denoted $\bar{\alpha}_{f,v}$. R_{vt}, a dummy for reallocation in each village-year, is then regressed on the mean return to lagged fertilizer, standardized to have mean zero and standard deviation one. A control for topographic characteristics T_v, also correlated with reallocation frequency, is also included.[18]

$$R_{vt} = \beta_1 \bar{\alpha}_{f,v} + \beta_2 T_v + \epsilon_{vt} \tag{1.17}$$

The objective of this regression is to identify whether there is a cross-sectional correlation between the returns to lagged investment and the probability of reallocation: more specifically, whether there are fewer reallocations when the returns to lagged investment are higher and thus the cost of a reallocation in terms of foregone agricultural investment is higher. The standard errors are calculated by bootstrapping the two-step procedure (the

[18]The topographic characteristic employed is the proportion of village land that is forested; this is one of the few topographic characteristics reported in the village survey and thus available at the village level. Geographic coordinates are available only at the county level.

estimation of the production function and the estimation of equation (1.17)) with clustering at the provincial level.

The results can be found in Table 1.7. In Panel A, the estimated coefficients capturing the correlation between returns to lagged investment estimated using OLS and the reallocation probability are significant and negative, consistent with the hypothesis that village officials respond to variation in the costs of reallocation. A one standard deviation increase in the returns to lagged fertilizer leads to a decline in the probability of reallocation of around 5 percentage points on a base probability of roughly 33%, a proportional effect of around 15%. These results are consistent across specifications in which the value of the grain harvest and value-added in grain production are employed as the dependent variable, outliers in returns to investment are trimmed, and the sample is restricted to only rice and wheat producers.

I also report in Panel A the coefficient on the measure of topographic variability T_v, which is negative as expected though noisily estimated: reallocations are most transactionally intensive, and thus less frequent, in areas with more challenging topography. While the same control variable is also included in all subsequent results estimating equation (1.17), the coefficient β_2 is not reported in Panels B through D for concision.

As a robustness check, the production function in equation (1.16) is re-estimated allowing for cross-sectionally varying lagged returns to agricultural labor and area sown, as well as fertilizer. The coefficient on lagged fertilizer and the mean coefficient on all lagged inputs are then employed as the independent variable in (1.17), and the results are shown in Panel B of Table 1.7. The coefficients are again negative and of comparable magnitude.

Panel A of Table 1.8 shows the results employing the return to agricultural inputs estimated using the dynamic panel methodology; details on the implementation of the bootstrap in this case can also be found in the appendix. The coefficients are generally of similar sign and magnitude, though the results employing grain value as the dependent variable in the production function are noisy. This suggests that the correlation between returns to lagged agricultural inputs and the frequency of reallocations does not reflect any systematic bias in the estimation of the production function.

Panel B of the same table shows one final robustness check in which the production function is re-estimated allowing the returns to inputs to vary at the level of the village v:

$$y_{it} = \sum_{j=1}^{J}(\alpha_j x_{ijt} + \alpha_f f_{ij,t-1}) + (\sum_{v=1}^{V} \nu_{vj}) \times (\sum_{j=1}^{J} \alpha_j x_{ijt}) + (\sum_{v=1}^{V} \nu_v) \times \alpha_f f_{ij,t-1} \tag{1.18}$$
$$+\mu_c + \nu_v + \gamma_t + \epsilon_{ivt}$$

This allows for the inclusion of province fixed effects in the estimation of equation (1.17), to test whether the correlation between returns to lagged fertilizer and reallocation behavior is evident within provinces; standard errors are clustered at the village level. The results can be found in Panel B of Table 1.8, and the coefficients are again negative, though smaller and not statistically significant using the two-step bootstrap. Even within provinces, villages with higher returns to fertilizer seem to report less frequent reallocations, consistent with the hypothesis that village leaders respond to the relative costs of reallocations in choosing whether or not to reallocate.

One potential challenge to this conclusion would be the possibility of reverse causation: villages with frequent reallocations exhibit lower returns to agricultural inputs precisely because they reallocate land frequently. There are two responses to this argument. First, the agricultural production function is estimated only using data from households with continuous tenure on the same plot last year and this year. Thus any direct effect of reallocations (transactional costs of swapping plots, for example) should not be a source of bias. Second, the reduced form results have already shown that areas with greater frequency of reallocations show lower levels of agricultural inputs. This should, all things equal, generate an upward bias in the estimated returns to inputs in areas with higher frequency of reallocations, a bias that runs in the opposite direction from the detected effect.

1.6.3 Difference-in-difference in crop composition

As an alternate strategy to address the endogeneity of the estimated returns to agricultural inputs, I employ a dif-in-dif specification that exploits differing climatic conditions conducive to the cultivation of different grain crops and differing price shocks across those crops. The interaction of climate and price shocks generates shifts in crop composition and thus shifts in the estimated returns to fertilizer. I can then test whether these estimated returns predict reallocation patterns.

First, I define a dummy variable equal to one if rice cultivation (as opposed to wheat, the other primary grain crop) is reported in a given village-year. Rice cultivation is more common in villages with higher mean precipitation, and more common in years where the reported rice price is higher. These relationships are shown in the first two columns of Table 1.9. The price employed is the price for mandatory grain quota sales to the government; all rural households are required to sell a certain inframarginal quantity of grain to the government at a price that is below market price, but varies across grains.[19] When the rice

[19]The rice quota price employed is constructed by calculating the reported quota price for each household

quota price increases, some villages who were previously not cultivating rice switch into rice production while villages already cultivating rice report no change. This generates a negative coefficient on the interaction of quota price and precipitation, shown in Column (3).

Using the dummy for rice cultivation, I then impute the lagged return to fertilizer using crop-specific returns estimated on the nationwide sample of data, again using both value added and grain revenue as the dependent variable. This estimated lagged return is denoted $\tilde{\alpha}_{f,v}$. For villages that report rice (no rice) production, $\tilde{\alpha}_{f,v}$ is set equal to the estimated returns to lagged fertilizer for rice (wheat). I then regress a dummy for reallocation on the estimated lagged return instrumented by the interaction of precipitation and price, including village and year fixed effects.

$$R_{vt} = \beta\tilde{\alpha}_{f,v} + \nu_v + \gamma_t \tag{1.19}$$

The key result shown in Columns (4) and (5) of Table 1.9 shows the expected negative correlation between the imputed return to lagged fertilizer and the probability of reallocation; Column (4) uses the estimated return to lagged fertilizer with grain revenue as the dependent variable, and Column (5) uses the estimated return from the value-added specification. This suggests that the negative correlation between medium-term returns to fertilizer and reallocation patterns does not simply reflect an underlying difference in unobservable characteristics across regions with varying frequencies of reported reallocations.

The exclusion restriction for this specification requires that an increase in the rice quota price has no disparate effect across areas with differing levels of precipitation other than via a shift in crop composition. I can control directly for a cross-sectionally varying effect of quota revenue, capturing the direct effect of a higher quota price for rice, and the estimated coefficient is not significantly different, though the standard error is slightly larger.

Taken together, the evidence of a correlation between reallocation propensity and the estimated returns to agricultural inputs is consistent with the hypothesis that village officials are selecting into reallocation on the basis of its relative costs in foregone agricultural investment. Thus despite the fact that reallocations and the associated tenure insecurity generate substantial costs, the decision by village leaders to implement them does not seem to reflect pure irrationality. This raises the question of what the benefits of reallocations are, for both officials and rural households, and whether the observed pattern of reallocations could in fact be optimal under certain conditions.

---

and calculating the mean quota price in each province and year for households that are solely rice producers.

1.7 Estimating the village leader's objective function

The conceptual model used to frame this analysis stated that village officials will choose to reallocate when the benefits exceed the costs, where the benefits were hypothesized to be greater intravillage equity and extraction of rents, and the costs were the loss in output induced by tenure insecurity and the transactional burden of conducting the reallocation. Given a parameterization of these elements, the observed pattern of reallocations can be used to infer the relative weights assigned to greater equity vis-a-vis foregone output and transactional costs in the village official's objective function using a revealed preference approach.

The benefits of reallocation B_{vt} are measured here by the increase in equity as a result of a reallocation. Reallocations on average do not result in a decrease in static measures of inequality in land distribution (e.g., the Gini coefficient). This presumably reflects the fact that the majority of land transfers implemented in reallocations are plot swaps, rather than reconfigurations of plots. Accordingly, households swap positions in roughly the same overall distribution of landownership.

For this reason, I employ a dynamic measure of inequality designed to capture the intuition that one of the primary objectives of reallocation is to ensure that households' average landholdings over time, relative to the size of the household, are (relatively) equitable. In other words, no household is characterized by landholdings per capita that are permanently above or below the median. First, I define $\tilde{L}_{it}$ as the within-household mean of land owned per capita for household i in periods t, t-1 and t-2. I then calculate three standard measures of inequality for this measure $\tilde{L}_{it}$, the Gini coefficient and the general entropy measures GE (1) and GE (2).

Table 1.10 shows the results from estimating the following equation in which inequality measures I_{vt} are regressed on a dummy for reallocation and village and year fixed effects.

$$I_{vt} = \beta_1 R_{vt} + \beta_2 I_{v,t-1} + \nu_v + \gamma_t + \epsilon_{vt} \tag{1.20}$$

The estimated coefficients suggest that reallocating land results in a significant decline in each measure of inequality, with the estimated magnitude of the effect between 5% and 10%.[20] Reallocations move households that previously had higher per-capita landholdings down in the landownership distribution and vice versa, thus generating convergence in the mean per-capita landholdings reported by each household over time. Moreover, reallocations

[20]This equation is estimated on the full panel of observed villages and years from 1987 to 2002.

are more common in years in which these measures of inequality are higher. This pattern is evident in Columns (3) through (6) of the same table, showing the results of the following regression:

$$R_{vt} = \beta I_{v,t-1} + \nu_v + \gamma_t + \epsilon_{vt} \tag{1.21}$$

The evidence suggests that increasing equity is one of the goals that village leaders seek to achieve when they reallocate land, and reallocations are at least a moderately effective tool in attaining this objective.

B_{vt} is then defined as the absolute value of the decline in the specified inequality measure (Gini or general entropy) induced by a reallocation, normalized on a scale of 0 to 100. In years that did not experience a reallocation, the counterfactual benefit $\hat{B}_{vt}$ is estimated as a random draw from a normal distribution that has the mean and standard deviation corresponding to the observed mean and standard deviation of reallocation benefit over all reallocations observed in that village.[21] T_v, the topographic burden of reallocation, is measured as the percentage of land in the village that is forested (again, on a scale of 0 to 100).

The cost C_{vt} of a reallocation is the estimated difference between output in the case of a reallocation in a given village and year and output in the absence of a reallocation. This difference is calculated employing the decline in sown area, labor and fertilizer predicted by the reduced form results, scaled by the mean risk of plot loss for households in a reallocation year.[22] C_{vt}, the total cost of a reallocation in village v in year t, is the sum of the difference in output across all H households observed in the village, valued at the market price in hundreds of yuan.

$$C_{vt} = \sum_{i=1}^{H} \Delta Y_{ivt} \tag{1.22}$$

The net benefit of reallocations ψ_{vt} is then defined as a simple quadratic function of the benefits and costs. The weight on the quadratic function of C, lost revenue due to decreased agricultural output, is normalized to one.

$$\psi_{vt} = \alpha_1(B_{vt} + B_{vt}^2) - \alpha_2(T_v + T_v^2) - (C_{vt} + C_{vt}^2) \tag{1.23}$$

If $\psi_{vt} \geq 0$, then a reallocation is optimal; if it is less than zero, a reallocation is not optimal. For postulated values of α_1 and α_2, a distribution of optimal reallocations can

[21] In estimation, I repeat this exercise using 100 random draws for each non-reallocation year.

[22] The decline in output is then calculated using the estimated returns to agricultural inputs allowing these returns to vary by province and crop, as in Section 1.6.2.

be generated and compared to the observed distribution of reallocations. The objective is to identify parameter values that best reproduce the observed pattern of reallocations. More specifically, I wish to identify parameters that maximize the accurate prediction rate across all (reallocation and non-reallocation) years, as well as minimizing the difference in prediction rates between reallocation and non-reallocation years.[23] Define π_T as the percent of all reallocation and non-reallocation events that are accurately predicted by the postulated parameters, and π_R and π_{NR} as the percent of reallocations and non-reallocations that are accurately predicted, calculated separately. The objective is to maximize $\hat{\pi}$, defined as

$$\hat{\pi} = \pi_T - \|\pi_R - \pi_{NR}\| \tag{1.24}$$

$\hat{\pi}$ is maximized by performing a grid search across potential values of α_1 and α_2. The range of parameters tested is 0 to 100 for both parameters; the increments of the grid are varied for each specification, and reported in the results table. Standard errors are bootstrapped across two hundred replications with re-sampling at the village-year level. For each specification, α_1 and α_2 are reported as well as π_R and π_{NR}.

Following this optimization process, I infer the predicted distribution of reallocations, conditional on the estimated weights, that would be optimal from the perspective of the official: namely, reallocating only when the net benefit is positive. I can then compare the estimated cost per reallocation of the optimal reallocations to the observed reallocations. The difference ΔC as a percentage of the cost of the observed reallocations is reported in the final row of Table 1.11.

The results show that the estimated weight on greater equity in the village leader's objective function is around 10, while the estimated weight on the transactional burden imposed by elevation is indistinguishable from zero. Converting these estimates to more easily understandable magnitudes, at the median level of C and B, officials are willing to trade off a 1% increase in equity in landownership against a 25% decline in revenue from grain production.

The estimated parameters predicts around 50% of the observed events, both reallocations and non-reallocations. Most importantly, comparing the implied distribution of optimal reallocations given these weights and the observed distribution, the foregone output as a result of reallocations would be around 40-50% lower per reallocation if village leaders reallocated

[23]If the objective was defined simply as maximizing π_T, given an observed reallocation rate of around one third, the resulting parameters predict the observed non-reallocations with a high degree of accuracy, while having little predictive power for reallocations.

only where the net benefit as estimated by ψ_{vt} were positive.

It is important to note that one potential reason that the estimated weight on equity in the village leader's objective function is relatively high could be that in fact there are substantial private benefits (e.g., rent extraction) for the official in conducting a reallocation that this analysis fails to take into account. However, given that households presumably place no value (or negative value) on these rents, this exercise is nonetheless informative about the relative weight on equity gains that would have to characterize the household's utility function in order for the observed pattern of reallocations to be optimal.

The results suggest first, that the observed distribution of reallocation decisions is consistent with village leaders placing a high weight on the benefit of greater equity compared to potential output losses. Second, even given this greater weight on equity, and despite the fact that village leaders are partially optimizing the choice of reallocations, they are also making significant and costly errors. Accordingly, the objective of enhanced equity could be achieved at considerably lower cost given a different set of reallocation decisions.

1.8 Conclusion

Although secure property rights are perceived as immensely important to economic development, the literature on the impact of inframarginal variation in property rights on economic outcomes remains relatively sparse. This paper contributes to this literature by evaluating one of the most unusual and far-reaching experiments in land property rights over the last half-century, the system of village-based reallocations of land in China. Implemented in order to maintain relative equity among households and to allow for adjustment of landholdings in absence of any rural land market, this system generates periodic disruptions in property rights for rural households, who have no guarantee that they will continue to farm the plot they currently hold.

Using an identification strategy that exploits intra-village variation in security of tenure, as well as cross-village variation in the propensity to reallocate land, this analysis finds that a lower probability of land reallocation has a substantial impact on households' economic behavior. Households that are less likely to see their tenure on their current plot disrupted by virtue of their past inclusion in a reallocation employ more agricultural inputs and produce more output than other households, and this effect is of substantial magnitude.

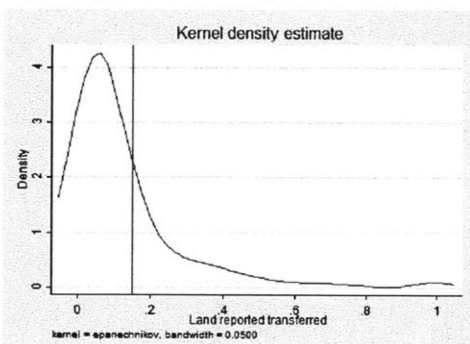
At the same time, there is evidence that officials respond to variation in the costs of disrupting property rights in choosing whether or not to hold a reallocation. Village leaders are less likely to reallocate in villages where disruptions to property rights are costly, but they appear to make some significant mistakes in reallocating where the net benefit, even given a substantial weight on greater equity in the official's objective function, is negative. Thus while property rights institutions at a micro-level adapt to reflect the relative costs and benefits of different institutional structures in different economic contexts, this adaptation process is far from perfect.

1.9 Figures and tables

Figure 1-1: Map of sample counties

Figure 1-2: Land transfers

(a) Proportion land reported transferred



(b) Prop. hh. reporting land changes

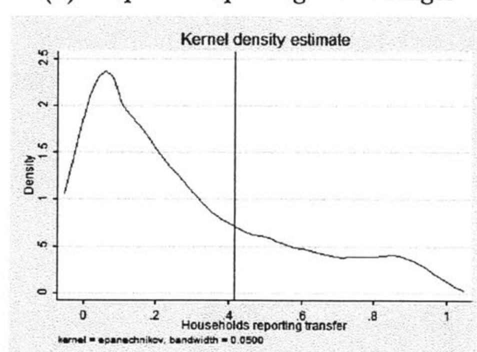


Figure 1-3: Probability of reallocation participation by decile of landownership

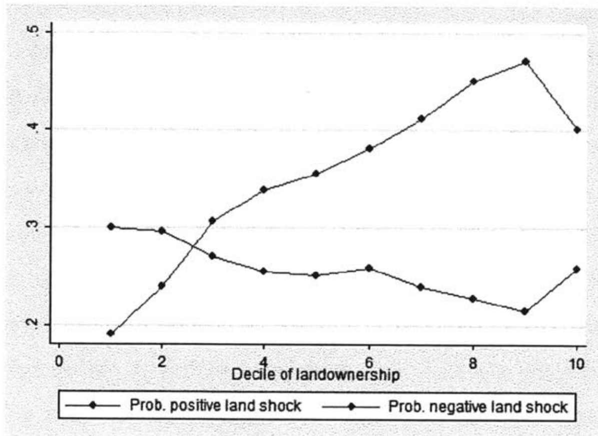
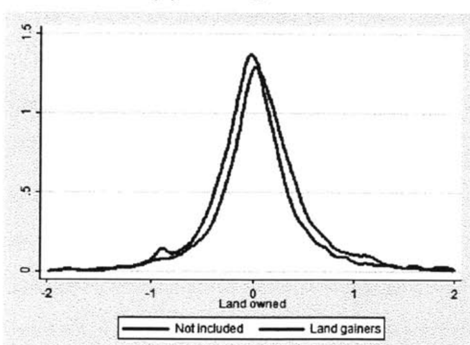


Figure 1-4: Kernel density estimates of landholding distributions

(a) Land gainers



(b) Land losers

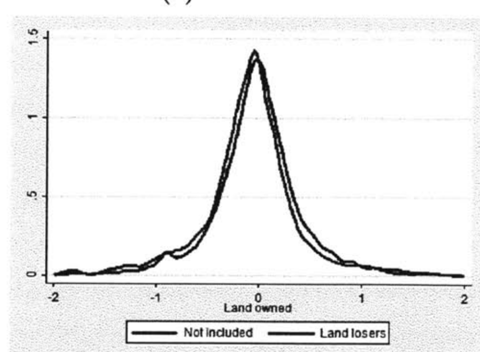
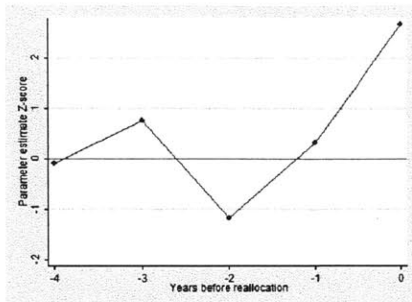
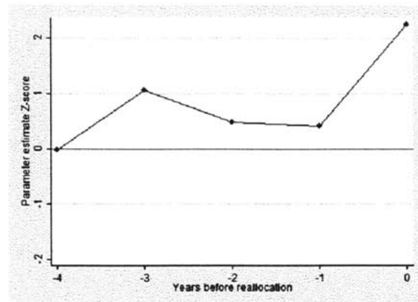


Figure 1-5: Anticipation of reallocation in pre-reallocation years

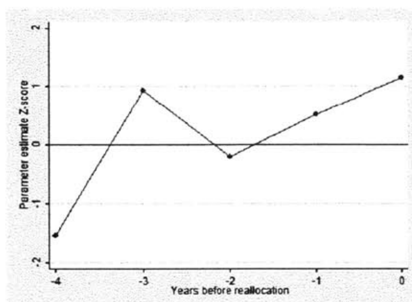
(a) Area sown



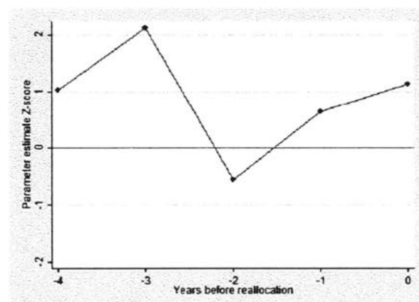
(b) Fertilizer



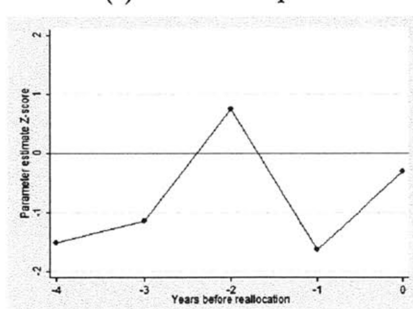
(c) Agricultural labor



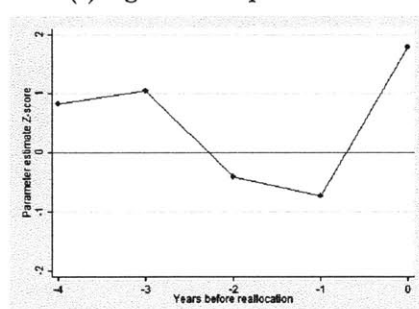
(d) Agricultural structures



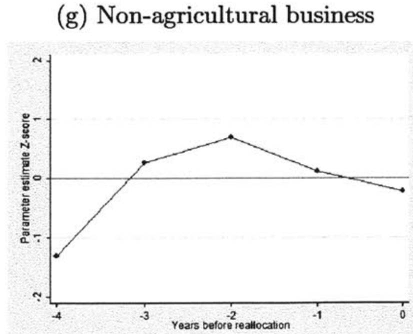
(e) Moveable capital



(f) Agricultural production



(g) Non-agricultural business



(h) Outside labor

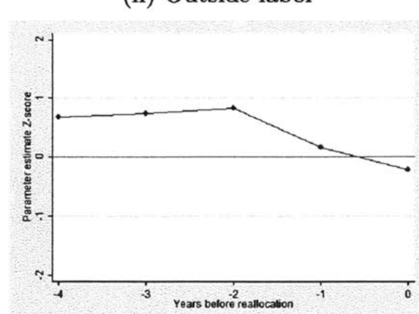


Figure 1-6: Reduced form coefficients for alternate definitions of reallocation

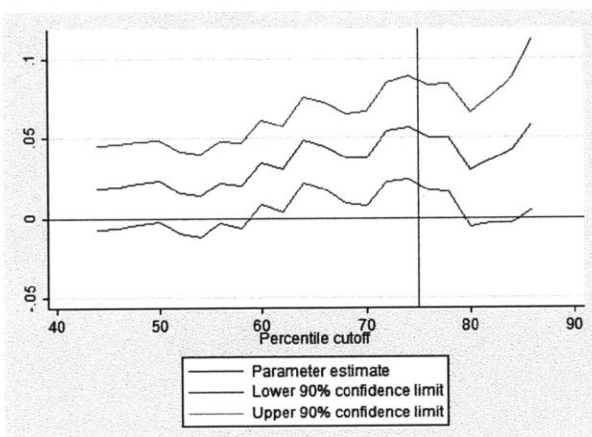


Table 1.1: Summary statistics

Village pop.	1690.794 (1134.75)
Village hh	417.90 (276.29)
Land per hh (hectare)	.40 (.37)
Plots per hh	5.96 (4.94)
Households sampled	68.42 (26.99)
Reallocation dummy	.53 (.50)
Forestry prop.	.22 (.29)

Table 1.2: Intravillage variation in reallocation probability: First stage

	D_{ivt} (1)	D_{ivt} (2)
DP_{ivt}^{-1}	-.038 (.014)***	
DN_{ivt}^{-1}	-.043 (.014)***	
D_{ivt}^{-1}		-.041 (.013)***
Mean D_{ivt}	.558	.558
F	5.436	10.831
Obs.	14973	14973

Notes: All specifications include village and year fixed effects, a control for vingtile of landownership, and standard errors clustered at the village-year level. The independent variables in Column (1) are dummy variables equal to one if a household's land was reallocated positively (DP_{ivt}^{-1}) or negatively (DN_{ivt}^{-1}) in the previous reallocation. The independent variable in Column (2) is a pooled dummy equal to one if the household had its land reallocated in a previous reallocation, D_{ivt}^{-1}. The sample is restricted to village-years in which a reallocation is observed. Asterisks denote significance at the 10, 5 and 1 percent level.

Table 1.3: Intravillage variation in reallocation probability: Reduced form

	Sown area (1)	Fertilizer Fertilizer (2)	Labor Labor (3)	Structures Structures (4)	Other cap. (5)	Agri. prod. (6)	Hh business (7)	Outside labor (8)
				Panel A: Reduced form				
$D_{ivt}^{-1} \times R_{vt}$	.067	.062	.041	.021	-.016	.062	-.005	-.011
	(.030)**	(.023)***	(.025)*	(.022)	(.023)	(.027)**	(.023)	(.026)
D_{ivt}^{-1}	.022	.013	.030	.007	.006	.016	-.00006	.031
	(.014)	(.011)	(.013)**	(.011)	(.009)	(.014)	(.013)	(.014)**
Obs.	46030	47841	46760	49376	49376	46465	49376	49376
				Panel B: Split reduced form				
$DP_{ivt}^{-1} \times R_{vt}$	.061	.042	.052	.031	-.012	.055	.016	.019
	(.036)*	(.024)*	(.029)*	(.026)	(.034)	(.032)*	(.030)	(.033)
$DN_{ivt}^{-1} \times R_{vt}$	.072	.074	.035	.015	-.018	.067	-.017	-.029
	(.034)**	(.026)***	(.028)	(.025)	(.025)	(.031)**	(.025)	(.029)
DP_{ivt}^{-1}	-.020	-.009	-.003	-.008	.015	-.020	-.002	.028
	(.015)	(.012)	(.015)	(.014)	(.013)	(.015)	(.015)	(.016)*
DN_{ivt}^{-1}	.047	.027	.051	.016	.00004	.038	.001	.032
	(.016)***	(.012)**	(.015)***	(.013)	(.011)	(.016)**	(.015)	(.016)**
Test: $\beta_1 = \beta_2$	.742	.165	.640	.651	.888	.702	.250	.139
Obs.	46030	47841	46760	49376	49376	46465	49376	49376

All specifications include village and year fixed effects interacted with R_{vt}, a control for vingtile of landownership also interacted with R_{vt}, and standard errors clustered at the village-year level. The independent variable in Panel A is the interaction between a pooled dummy equal to one if the household had its land reallocated in a previous reallocation, D_{ivt}^{-1}, and a dummy for a current reallocation R_{vt}. The independent variables in Panel B are the interactions between dummy variables equal to one if a household's land was reallocated positively (DP_{ivt}^{-1}) or negatively (DN_{ivt}^{-1}) in the previous reallocation and R_{vt}. The dependent variables are sown area, fertilizer, agricultural labor, agricultural structures, tools and animals owned, and dummies for participating in a non-agricultural business or in outside labor; sown area, fertilizer, agricultural production and agricultural labor are reported per acre owned, and all variables are normalized relative to the control observations. Asterisks denote significance at the 10, 5 and 1 percent level. The final row of Panel B reports the p-value for a test of equality on the coefficients of $DP_{ivt}^{-1} \times R_{vt}$ and $DN_{ivt}^{-1} \times R_{vt}$.

Table 1.4: Intravillage variation in reallocation probability: 2SLS estimates

	Sown area (1)	Fertilizer Fertilizer (2)	Labor Labor (3)	Structures Structures (4)	Other cap. (5)	Agri. prod. (6)	Hh business (7)	Outside labor (8)
				Panel A: IV estimates				
Allocation dummy	-1.581	-1.347	-1.059	-.500	.376	-1.433	.114	.259
	(.829)*	(.633)**	(.742)	(.546)	(.544)	(.733)*	(.557)	(.621)
Obs.	46030	47841	46760	49376	49376	46465	49376	49376
				Panel B: IV with polynomial in land area				
Allocation dummy	-1.604	-1.325	-1.055	-.457	.387	-1.410	.107	.243
	(.794)**	(.609)**	(.708)	(.525)	(.514)	(.693)**	(.535)	(.597)
Obs.	46030	47841	46760	49376	49376	46465	49376	49376
				Panel C: IV excluding households with past demographic instability				
Allocation dummy	-2.993	-2.428	-2.580	-.612	.003	-1.916	-.896	.145
	(1.525)**	(1.139)**	(1.602)	(.866)	(.956)	(1.272)	(.900)	(.963)
Obs.	22724	23493	23177	24324	24324	22912	24324	24324
				Panel D: IV excluding shrinking households				
Allocation dummy	-1.565	-1.388	-1.225	-.282	.565	-1.346	-.088	.280
	(.877)*	(.726)*	(.827)	(.567)	(.632)	(.826)	(.616)	(.666)
Obs.	40461	42108	41138	43514	43514	40855	43514	43514
				Panel E: IV for households below median of landownership				
Allocation dummy	-1.530	-1.291	-1.077	-.828	.220	-1.383	-.464	-.142
	(.870)*	(.618)**	(.894)	(.585)	(.419)	(.818)*	(.568)	(.637)
Obs.	22336	23510	22530	24877	24877	22554	24877	24877

Notes: All specifications include village and year fixed effects interacted with R_{vt}, a control for vingtile of landownership also interacted with R_{vt}, a control for D_{ivt}^{-1} and standard errors clustered at the village-year level. The independent variable is a dummy for a household having its land reallocated, instrumented by $D_{ivt}^{-1} \times R_{vt}$. The dependent variables are sown area, fertilizer, agricultural labor, agricultural structures, tools and animals owned, and dummies for participating in a non-agricultural business or in outside labor; sown area, fertilizer, agricultural production and agricultural labor are reported per acre owned, and all variables are normalized relative to the control observations. Asterisks denote significance at the 10, 5 and 1 percent level. In Panel B, a quadratic polynomial in land area is added. In Panel C, the sample is restricted to households with no past history of demographic shifts in reallocation years; in Panel D, the sample is restricted to households that report either constant or increasing household size; in Panel E, it is restricted to households in the lowest five deciles of landownership.

Table 1.5: Information as a channel for predicting reallocations

	Allocation pos. (1)	Allocation neg. (2)		Allocation pos. (3)	Allocation neg. (4)
Rice int.	.080 (.065)	-.092 (.078)	Retail int.	.016 (.043)	-.039 (.043)
Wheat int.	.036 (.062)	-.227 (.082)***	Fish int.	.052 (.061)	.086 (.058)
Husb. int.	.080 (.037)**	-.103 (.038)***	Educ. int.	-.039 (.033)	.132 (.034)***
Manu. int.	-.022 (.052)	.026 (.059)	Vet int.	-.015 (.121)	.049 (.069)
Trans. int.	-.086 (.087)	.158 (.102)	Grandparent int.	-.030 (.023)	.059 (.029)**
Cons. int.	-.115 (.166)	-.157 (.165)	Party int.	.005 (.031)	-9.02e-06 (.038)

Notes: Each cell corresponds to a separate regression including village and year fixed effects and a control for vingtile of landownership; standard errors are clustered at the village-year level. The dependent variable is a dummy for positive or negative changes in land in a reallocation as indicated; the independent variable reported is the interaction between a household dummy of interest and the mean of that dummy among government officials' households in that village-year. The dummy variables are indicators for whether the household engages in rice or wheat cultivation, or husbandry, manufacturing, transportation, construction, retail or fishing as a household business, as well as indicators for the presence within the household of a principal laborer with education beyond high school, a veteran of the armed forces, residential grandparents, or a member of the Communist party. Additional independent variables not reported are the household and official dummy entering linearly. The sample is restricted to years in which a reallocation occurs. Asterisks denote significance at the 10, 5 and 1 percent level.

Table 1.6: Estimated returns to agricultural inputs

	OLS				Dynamic panel GMM			
	Full sample		Rice and wheat prod.		Full sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grain area	.620 (.022)***	.607 (.025)***	.576 (.029)***	.631 (.018)***	0.667 (0.192)***	0.664 (0.184)***	0.647 (0.178)***	0.611 (0.189)***
Labor	.126 (.013)***	.134 (.014)***	.130 (.019)***	.125 (.012)***	0.035 (0.193)	0.133 (0.185)	0.108 (0.180)	0.234 (0.191)
Fertilizer	.120 (.009)***	.118 (.010)***	.131 (.015)***	.102 (.009)***	0.324 (0.151)**	0.288 (0.102)***	0.308 (0.118)***	0.264 (0.106)**
Lagged fertilizer		.015 (.006)***		.014 (.006)**			0.122 (0.094)	0.158 (0.118)
Lags as instruments					Full set	Lags 3-4	Full set	Lags 3-4
Sargan-Hansen p-value					0.017	0.011	0.195	0.095
Test of overidentifying restrictions					0.824	0.616	0.875	0.912
Obs.	51073	42460	33789	28383	30019	30019	30019	30019

Notes: Each column represents a separate regression estimating the returns to agricultural inputs; the dependent variable is value added in grain production (the grain harvest valued at the market price in each village-year minus the cost of seeds). Columns 1 through 4 report estimates of the returns to agricultural inputs estimated in an OLS specification with village, year and crop fixed effects; in columns 3 and 4, the specification is restricted to rice and wheat producers. Columns 5 through 8 report estimation results employing a dynamic panel GMM methodology, employing the full sample. Asterisks denote significance at the 10, 5 and 1 percent level.

Table 1.7: Returns to lagged fertilizer and reallocation probability

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Lagged fertilizer				
Returns to lagged fertilizer	-.097	-.092	-.090	-.084	-.069	-.038	-.080	-.073
	[.030]***	[.033]***	[.037]**	[.043]**	[.025]**	[.032]	[.032]**	[.042]*
Forestry prop.	-.084	-.086	-.088	-.079	-.109	-.102	-.081	-.091
	[.056]	[.058]	[.060]	[.059]	[.072]	[.082]	[.057]	[.059]
Prod. fun. sample	All households				Rice and wheat producers			
Prod fun. dep. variable	Value added		Grain prod.		Value added		Grain prod.	
Outliers trimmed	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	896	847	811	733	896	811	811	761
				Panel B: Lagged fertilizer, area and labor				
Returns to lagged fertilizer	-.106		-.043		-.088		-.058	
	[.036]**		[.035]		[.038]**		[.031]*	
Mean return to lagged inputs		-.058		-.045		-.068		-.046
		[.031]*		[.041]		[.034]**		[.044]
Prod. fun. sample	All households				Rice and wheat producers			
Prod fun. dep. variable	Value added		Grain prod.		Value added		Grain prod.	
Obs.	896	896	896	896	811	811	811	811

Notes: The dependent variable is reallocation at the village-year level; all standard errors are clustered at the province level, and the independent variables are standardized to have mean zero and standard deviation one. The independent variable is the mean estimated return to lagged fertilizer or to other lagged inputs in the village-year, normalized by the standard error. Returns to agricultural inputs are allowed to vary by province and crop; all households and rice/wheat producers denotes the sample used to estimate the production function, and value-added or grain value denotes the dependent variable used. All specifications include a control for topographic variability (the proportion of land forested), though the coefficient is reported only in Panel A for concision. Asterisks denote significance at the 10, 5 and 1 percent level.

Table 1.8: Returns to lagged fertilizer and reallocation probability

Panel A: Dynamic panel coefficients

Returns to lagged fertilizer	-.072 [.045]	.023 [.041]	-.082 [.037]**	-.034 [.045]
Prod. fun. sample	All households		Rice and wheat producers	
Prod fun. dep. variable	Value added	Grain prod.	Value added	Grain prod.
Obs.	848	644	816	585

Panel B: Lagged fertilizer with variation at the village level

Returns to lagged fertilizer	-.023 [.020]	-.021 [.018]	-.030 [.021]	-.030 [.028]	-.035 [.024]	-.033 [.021]	-.030 [.023]	-.020 [.029]
Prod. fun. sample	All households				Rice and wheat producers			
Prod fun. dep. variable	Value added		Grain prod.		Value added		Grain prod.	
Fixed effect	Prov.	Prov.	Prov.	Prov.	Prov.	Prov.	Prov.	Prov.
Outliers trimmed	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	853	845	844	836	837	831	818	812

Notes: The dependent variable is reallocation at the village-year level; all standard errors are clustered at the province level, and the independent variables are standardized to have mean zero and standard deviation one. The independent variable is the mean estimated return to lagged fertilizer in the village-year, normalized by the standard error. In Panel A, the coefficients in the agricultural production function are estimated using dynamic panel GMM. In Panel B, returns to agricultural inputs are allowed to vary by village. All specifications include a control for topographic variability (the proportion of land forested), though the coefficient is not reported for concision. Asterisks denote significance at the 10, 5 and 1 percent level.

Table 1.9: Difference-in-difference estimates of lagged fertilizer returns

	Rice dummy			Reallocation	
	(1)	(2)	(3)	(4)	(5)
Precipitation	.272 (.029)***				
Rice quota price		.020 (.004)***			
Precipitation price int			-.157 (.063)**		
Fertilizer return				-2.521 (1.508)*	-.776 (.464)*
Obs.	903	1077	847	847	847
F	88.733	17.329	5.1	2.277	2.277

Notes: The dependent variable is Columns (1) to (3) is a dummy for whether rice is cultivated in a given village-year. The dependent variable in Columns (5) and (6) is a dummy for whether a reallocation occurs. Mean precipitation, the rice quota price and the interaction are normalized to have mean zero and standard deviation one. The estimated return to lagged fertilizer is imputed using the dummy for rice cultivation and the returns to lagged fertilizer for rice and wheat production estimated on the full sample. Asterisks denote significance at the 10, 5 and 1 percent level.

Table 1.10: Benefits of reallocation

	Gini	GE 1	GE 2		Reallocation	
	(1)	(2)	(3)	(4)	(5)	(6)
Reallocation	-.007 (.001)***	-.004 (.001)***	-.005 (.002)***			
Lagged Gini	.770 (.017)***			.933 (.321)***		
Lagged GE(1)		.857 (.018)***			.947 (.368)**	
Lagged GE(2)			.842 (.020)***			.570 (.290)**
Mean dep. variable	.182	.066	.070	.352	.352	.352
Obs.	1664	1664	1664	1664	1664	1664

Notes: The dependent variable is Columns (1) to (3) is the specified measure of inequality in $\tilde{L}_{it}$, defined as the mean of per-capita landholdings for household i over period t and the two preceding periods. The dependent variable in Columns (4) through (6) is a dummy for reallocation at the village-year level. All regressions include village and year fixed effects; asterisks denote significance at the 10, 5 and 1 percent level.

Table 1.11: Parameters of the village leader's objective function

	(1)	(2)	(3)	(4)	(5)	(6)
α_1	6	6	12	12	10	10
	(14.044)	(36.216)	(2.100)***	(2.302)***	(3.061)***	(15.315)
α_2	0	0	0	0	0	0
	(.171)	(.242)	(0)	(0)	(0)	(.086)
π_R	0.504	0.504	0.444	0.444	0.452	0.452
	(.026)***	(.028)***	(.034)***	(.031)***	(.026)***	(.024)***
π_{NR}	0.495	0.495	0.446	0.446	0.455	0.455
	(.025)***	(.025)***	(.035)***	(.032)***	(.029)***	(.029)***
ΔC	-0.498	-0.498	-0.459	-0.459	-0.417	-0.417
	(.243)**	(.550)	(.130)**	(.128)**	(.134)**	(.203)**
Inequality measure employed	Gini	Gini	GE(1)	GE(1)	GE(2)	GE(2)
Grid step (α_1)	1	1	1	1	1	1
Grid step (α_2)	1	.5	1	.5	1	.5

Notes: The coefficients correspond to the estimated weights on equity and transactional costs in the village leader's objective function. π_R and π_{NR} are the proportion of reallocation and non-reallocation events respectively predicted by the estimated parameters. ΔC is the percent difference in cost between the observed distribution of reallocations and the optimal distribution of reallocations conditional on the estimated weights. Standard errors are calculated using a bootstrap with 200 replications.

1.A Appendix: Household optimization problem

Assume the household seeks to maximize value-added profits in agricultural production (i.e., profits minus the cost of seeds); the production function is not constrained to be constant returns to scale, and evidence suggests it is in fact decreasing returns to scale. I postulate a standard Cobb-Douglas production function in which there are lagged returns to investment (fertilizer). Note that fertilizer is assumed to be a flow variable: F_t is equal to fertilizer applied in period t only. However, fertilizer applied in period t-1 is allowed to continue to have a direct effect on soil productivity.

γ is equal to the probability of reallocation, identical in every period; in the case of a reallocation, lagged returns to fertilizer are lost. Accordingly, the production function and value-added profits take the following form. The contemporaneous return to fertilizer will be denoted α_C, and the lagged return to fertilizer denoted α_F.[24]

$$
\begin{aligned}
Y_t &= \tilde{A}_t L_t^{\alpha_L} N_t^{\alpha_N} F_t^{\alpha_C}(1-\gamma) F_{t-1}^{\alpha_F} && (1.25)\\
\pi_t &= P_t Y_t - P_t^s S_t && (1.26)
\end{aligned}
$$

Assume further that the household optimally chooses F_t and N_t, fertilizer and labor inputs, and that land cultivated L_t is a mechanical function of inputs chosen: i.e., when a household optimally uses more inputs, it will cultivate more of its land allotment. I will focus on analyzing the household's optimization problem in period t, assuming there was no reallocation in the last period (t-1). For simplicity of notation, in the subsequent analysis denote $A_t = \tilde{A}_t L_t^{\alpha_L} F_{t-1}^{\alpha_F}$.[25]

Define σ_t as the return next period to this period's investment in the absence of a reallocation.

$$
\begin{aligned}
\sigma_t &= \frac{\partial \pi_{t+1}}{\partial F_t} && (1.27)\\
&= \alpha_F \tilde{A}_{t+1} L_{t+1}^{\alpha_L} F_t^{\alpha_F - 1} F_{t+1}^{\alpha_C} N_{t+1}^{\alpha_N} && (1.28)
\end{aligned}
$$

The first-order condition governing optimal fertilizer and labor can then be written as follows.

[24] This model could easily be generalized to allow for lagged returns to all agricultural inputs. For simplicity, I focus here on the exposition for the case in which only fertilizer has lagged returns.

[25] If there was a reallocation in period t-1, then $A_t = \tilde{A}_t L_t^{\alpha_L}$. This assumption does not affect the analysis that follows.

$$\frac{w_t}{r_t} = \frac{P_t\alpha_N A_t F_t^{\alpha_C} N_t^{\alpha_N-1}}{P_t\alpha_C A_t F_t^{\alpha_C-1} N_t^{\alpha_N} + P_{t+1}\sigma_t(1-\gamma)} \tag{1.29}$$

$$0 = \frac{P_t\alpha_N A_t F_t^{\alpha_C} N_t^{\alpha_N-1}}{P_t\alpha_C A_t F_t^{\alpha_C-1} N_t^{\alpha_N} + P_{t+1}\sigma_t(1-\gamma)} - \frac{w_t}{r_t} \tag{1.30}$$

Denoting the right-hand side of (1.30) by ψ, the implicit derivative of fertilizer investment with respect to the probability of reallocation can be calculated as $\frac{\partial F_t}{\partial\gamma} = -\frac{\psi'(\gamma)}{\psi'(F_t)}$. Again for ease of notation, define λ_1 and λ_2:

$$\lambda_1 = P_t\alpha_N A_t F_t^{\alpha_C} N_t^{\alpha_N-1} \tag{1.31}$$

$$\lambda_2 = P_t\alpha_C A_t F_t^{\alpha_C-1} N_t^{\alpha_N} + P_{t+1}\sigma_t(1-\gamma) \tag{1.32}$$

The implicit derivative can then be calculated employing the following formula:

$$\begin{aligned}
\psi'(\gamma) &= \frac{\lambda_1}{(P_t\alpha_C A_t F_t^{\alpha_C-1} N_t^{\alpha_N} + P_{t+1}\sigma_t(1-\gamma))^2} \\
\psi'(F) &= \frac{1}{\lambda_2^2}(\lambda_2 P_t\alpha_N\alpha_C A_t F_t^{\alpha_C-1} N_t^{\alpha_N-1} \\
&\quad - \lambda_1 P_t\alpha_C(\alpha_C-1)A_t F_t^{\alpha_C-2} N_t^{\alpha_N} \\
&\quad - \lambda_1 P_{t+1}(1-\gamma)A_{t+1}\alpha_P(\alpha_F-1)F_t^{\alpha_F-2} F_{t+1}^{\alpha_C} N_{t+1}^{\alpha_N})
\end{aligned} \tag{1.33}$$

Given that both the numerator and denominator are positive, the implicit derivative formula yields that $F'(\gamma) < 0$. This is intuitive: optimal fertilizer investment declines when the probability of reallocation increases.

Assuming that households know whether or not they will lose their plot at the end of the year at the time they make their investments (i.e., γ is a dummy variable equal to zero or one), optimal fertilizer investment with or without a reallocation can be specified as follows. $F_t^R = \frac{w_t\alpha_C}{r_t\alpha_N}N_t$, while F_t^{NR} solves the following equation:

$$0 = \frac{P_t\alpha_N A_t F_t^{\alpha_C} N_t^{\alpha_N-1}}{P_t\alpha_C A_t F_t^{\alpha_C-1} N_t^{\alpha_N} + P_{t+1}\sigma} - \frac{w_t}{r_t} \tag{1.34}$$

Define the difference in investment between the reallocation and the non-reallocation case as follows.

$$\Delta F_t \equiv F_t^{NR} - F_t^R \tag{1.35}$$

ΔF_t is increasing in α_F, a comparative static that can be established again using the formula for implicit differentiation. Note that F_t^R is independent of the returns to lagged investment, while $\frac{\partial F_t^{NR}}{\partial \alpha_F} = -\frac{\psi'(\alpha_F)}{\psi'(F_t^{NR})}$. The denominator is positive, and the numerator can be written as follows:

$$\psi'(\alpha_F) = -\frac{\lambda_1}{\lambda_2^2} P_{t+1}[A_{t+1} F_t^{\alpha_F - 1} F_{t+1}^{\alpha_C} N_{t+1}^{\alpha_N} + A_{t+1} \alpha_P F_{t+1}^{\alpha_C} N_{t+1}^{\alpha_N} F_t^{\alpha_F - 1} (log F_t)] \tag{1.36}$$

Accordingly, $\frac{\partial F_t^{NR}}{\partial \alpha_F} > 0$ and thus $\frac{\delta \Delta F_t}{\partial \alpha_F} > 0$. The investment gap between years with and without reallocation is increasing in the returns to lagged fertilizer investment.

1.B Appendix: Dynamic panel estimation of the agricultural production function

Following Blundell & Bond (2000), an AR(1) error structure is now imposed on the production function.

$$\begin{aligned} y_{it} &= \alpha_l l_{it} + \alpha_s s_{it} + \alpha_n n_{it} + \alpha_f f_{it} + \alpha_p f_{i,t-1} + \gamma_t + (\eta_i + \nu_{it} + m_{it}) && (1.37) \\ v_{it} &= \rho v_{i,t-1} + e_{it} && (1.38) \end{aligned}$$

This model has a dynamic representation:

$$\begin{aligned} y_{it} = {} & \alpha_l l_{it} - \rho\alpha_l l_{i,t-1} + \alpha_s s_{it} - \rho\alpha_s s_{i,t-1} + \alpha_n n_{it} - \rho\alpha_n n_{i,t-1} \\ & + \alpha_f f_{it} + (\alpha_p - \rho\beta_f) f_{i,t-1} - \rho\beta_p f_{i,t-2} + \rho y_{i,t-1} \\ & + (\eta_i(1-\rho) + e_{it} + m_{it} - \rho m_{i,t-1}) \end{aligned} \tag{1.39}$$

The dynamic model can be rewritten as follows:

$$\begin{aligned} y_{it} = {} & \pi_1 l_{it} + \pi_2 l_{i,t-1} + \pi_3 s_{it} + \pi_4 s_{i,t-1} + \pi_5 n_{it} + \pi_6 n_{i,t-1} \\ & + \pi_7 f_{it} + \pi_8 f_{i,t-1} + \pi_9 f_{i,t-2} + \pi_{10} y_{i,t-1} + \gamma_t^* + (\eta_i^* + w_{it}) \end{aligned} \tag{1.40}$$

subject to the following non-linear common factor restrictions,

$$\begin{aligned} \pi_1 &= -\pi_2/\pi_{10} && (1.41) \\ \pi_3 &= -\pi_4/\pi_{10} && (1.42) \\ \pi_5 &= -\pi_6/\pi_{10} && (1.43) \\ \pi_7 &= -\pi_8/\pi_{10} - \pi_9/\pi_{10}^2 && (1.44) \end{aligned}$$

as well as equalities in π_1, π_3, π_5, π_7, π_8 and π_{10}.

Given consistent estimates of the unrestricted parameter vector π and $var(\pi)$, these restrictions can be tested and imposed using a minimum distance model to obtain the restricted parameter vector.

1.B.1 Estimating the unrestricted parameter vector

The unrestricted parameter vector is estimated using dynamic panel methods; the following exposition largely follows Blundell & Bond (2000). A standard assumption on the initial conditions ($E[x_{i1}e_{it}] = E[x_{i1}m_{it}] = 0$ for $t = 2 \ldots T$) yields the following moment conditions.

$$E[x_{i,t-s}, \Delta w_{it}] = 0 \tag{1.45}$$

for $s \geq 3$ where $w_{it} \sim MA(1)$. This allows for the use of lagged levels of the variables as instruments after the equation is first-differenced.

However, the resulting GMM estimator in first differences can have poor finite sample properties when the instruments (lagged levels) are weak. Imposing additional conditions on the correlation between the fixed effect and first-differenced variables allows for the generation of additional moment conditions that can be used to estimate the parameters. The additional assumptions needed are as follows:

$$E[\Delta x_{it}\eta_i^*] = 0 \tag{1.46}$$

$$E[\Delta y_{it}\eta_i^*] = 0 \tag{1.47}$$

The moment conditions thus implied can be written as follows, for $s = 2$ when $w_{it} \sim MA(1)$.

$$E[\Delta x_{i,t-s}(\eta_i^* + w_{it})] = 0 \tag{1.48}$$

In other words, lagged first differences of the variables can be used as instruments in the equations in levels. Both sets of moment conditions can be employed in a linear GMM estimator using both first-differenced and levels equations; this is what Blundell-Bond deem the system GMM estimator.

1.B.2 Estimating the minimum distance model

The minimum distance model entails minimizing the distance between the unrestricted parameter vector and the previously enumerated set of common factor restrictions $g(\hat{\pi})$.

$$f(\beta, g(\hat{\pi})) = H\beta - g(\hat{\pi}) = 0 \tag{1.49}$$

where $g(\hat{\pi})$ can be written as

$$\begin{bmatrix} \pi_1 \\ -\pi_2/\pi_{10} \\ \pi_3 \\ -\pi_4/\pi_{10} \\ \pi_5 - \pi_6/\pi_{10} \\ \pi_7 \\ -\pi_8/\pi_{10} - \pi_9/\pi_{10}^2 \\ \pi_8 \\ \pi_{10} \end{bmatrix}$$

The minimum distance estimator is given by the minimization of

$$D(\pi) = f(\beta, g(\hat{\pi}))'\hat{V}[g(\hat{\pi})]^{-1}f(\beta, g(\hat{\pi})) \tag{1.50}$$

where $\hat{V}[g(\hat{\pi})]$ denotes the estimated variance-covariance matrix of $g(\hat{\pi})$, estimated using the delta method. Minimization of D yields the following:

$$\hat{\beta} = (H'\ \hat{V}[g(\hat{\pi})]^{-1}\ H)^{-1}\ H'\ \hat{V}[g(\hat{\pi})]^{-1}\ g(\hat{\pi}) \tag{1.51}$$

with variance-covariance matrix

$$\hat{V}[\hat{\beta}] = (H'\ \hat{V}[g(\hat{\pi})]^{-1}\ H)^{-1} \tag{1.52}$$

1.B.3 Estimating by province and crop

Estimating and imposing the minimum distance restrictions in an equation including interactions between the primary agricultural inputs and province and crop dummies imposes too large a computational burden. Accordingly, in the results restricted to rice and wheat producers, the model is estimated separately for each province-crop pair provided there are adequate observations. In the full-sample specification, the results are estimated for each crop-province pair for which there are adequate observations, and then for the remaining pool of households in that province.

The bootstrap is implemented by bootstrapping with replacement at the household level for each province-crop, estimating the agricultural production function for each province-crop, and then estimating the mean return to lagged fertilizer in each village-year. 100 replications are employed.

Chapter 2

Long-Run Impacts of Land Regulation: Evidence from Tenancy Reform in India

2.1 Introduction

The institutional arrangements that shape access to land are central to the functioning of an agricultural economy. Given that a large fraction of the world's poor remain dependent on agriculture, production relations in this sector have a first-order impact on aggregate poverty. Moreover, a classic view formalized in Matsuyama (1992) contends that productivity improvements in agriculture are a spur to industrial development because they increase the demand for industrial goods. Accordingly, promoting the transfer of land to its highest-return use has the potential to have a major impact on economic growth in developing countries.

Institutions introduced or strengthened by colonial powers left a legacy of significant concentration of land ownership and insecure tenure for tenants in much of the developing world. In conjunction with imperfections in other key markets (e.g., the market for credit), these inequalities continue to constrain long-run economic growth and, in particular, the transfer of land towards high return activities.[1] As increasing demand for land from the industrial sector has clashed with often inefficient inherited institutions in many countries,

[1]See for example Pande & Udry (2006), Banerjee & Iyer (2005), Acemoglu, Johnson & Robinson (2001) and Binswanger, Deininger & Feder (1995). Banerjee (2003) provides an overview of the importance of credit market imperfections in development.

most notably India and China, intense debates have emerged around whether and how governments should regulate the terms on which land can be acquired from landowners and tenants (Ghatak & Mookherjee 2011).

Given the potential mismatch between relatively stagnant institutions and rapidly evolving demand, government regulation of land transactions is, unsurprisingly, common. However, there is no guarantee that such regulation produces gains for all groups even if there is greater overall efficiency. The complexity of designing partial reform in a second-best world is a longstanding theme in the development literature (Stiglitz 1988). But in large part due to data constraints, there is little solid empirical evidence on the long-run impact of regulated land markets. This, in turn, limits our understanding of whether and how economic actors use land markets to reduce or amplify the intended impact of regulation.

This paper explores the long-run effects of tenancy reforms using a unique natural experiment in India which, following Independence, witnessed a wave of state-level reforms (Appu 1996). The major period of reform in the four Southern Indian states examined here (Andhra Pradesh, Karnataka, Kerala and Tamil Nadu) began shortly after Independence and continued until the early 1970s. We employ village- and household-level data to trace the impact of reforms which took place more than thirty years prior to our survey, allowing us to examine a number of dimensions in which we could expect tenancy reform to have a long-run impact.

Theoretically, landlords can choose between different ways of exploiting their land to generate a return, including selling the land and investing in other assets. The attractiveness of operating land when tenants have stronger rights depends on the extent to which landlords can extract returns from doing so, while the ability to sell land depends on the capital market opportunities of potential owner-cultivators. Tenancy reforms lower the returns of renting land for landlords; thus it is logical to expect less use of tenancy and more land sales, particularly to those with access to the credit market. This will lead, in turn, to a change in the distribution of land ownership. If frictions in the land market allow landowners to extract only part of the surplus created in a land sale, sales will occur only to relatively high productivity individuals. Thus, by enabling more efficient land use, land sales will increase labor demand and hence agricultural wage.

Tracing through these equilibrium effects complicates the overall welfare impact. Cultivators who remain as tenants will gain, but marginal tenants will lose out as they become landless laborers. However, their opportunities in the labor market should improve. Households with better capital market opportunities are more likely to end up as owner-cultivators.

These are the predictions that we bring to the data.

Our identification strategy exploits the 1956 reorganization of state boundaries in Southern India, designed to transform the state units inherited from the British into linguistically coherent states. The reorganization allocated sub-district administrative units called blocks to states, on the basis of the population's linguistic composition. However, the need to form states with contiguous territory sometimes led to blocks with similar linguistic and cultural characteristics being assigned to different neighboring states. These blocks were analogous both in historical experience and social structure – two factors which, as we describe in Section 2.2, were significant determinants of land structure. The blocks, however, subsequently experienced significantly different programs of land reform.

We undertook a multi-stage sampling and survey procedure to construct our sample. We identified nine neighboring district pairs in the four Southern states. For each pair, blocks were matched using a linguistic index based on census data on the proportion population speaking each one of the eighteen languages reported spoken in the region. The eighteen best matched pairs were chosen, and in 2002 we conducted household surveys in a random sample of 259 villages spread across these blocks. In Section 2.4, we use 1951 data on village-level landlessness to demonstrate similarity in initial landlessness across blocks in matched pairs.

Our analysis, therefore, exploits variation in land reform across block pairs matched on linguistic characteristics to evaluate the impact of land reform. Accordingly, the key identifying assumptions require that the assignment of different blocks to different states along the border is quasi-random conditional on observable characteristics, and that the channel through which state assignment affects rural land distribution is land reform. If these assumptions hold, estimating the impact of land reform within these block pairs allows for an unbiased estimate of the impact of land reform on economic outcomes.

In addition, we interact variation in land reform with households' presumed land ownership prior to the reform, proxied by their caste status. This interaction both tests the key theoretical predictions about the differential impact of land reform on households with different baseline characteristics, and allows for the estimation of causal effects of land reform under the weaker identification assumption of no systematic variation in between-caste group differences across state borders.

The results suggest that tenancy reform reduced land inequality within villages, predominantly by transferring land from upper caste landowners to middle caste tenants. However, in line with the theory, tenancy reform also increased the number of landless Scheduled Caste and Scheduled Tribe (SC/ST) households, a group that presumably had poorer access to credit. Consistent with our model, we also observe higher agricultural wage after tenancy reform.

Our findings contribute to a large literature on institutional persistence (see for example ?). While the relationship between institutional patterns and economic outcomes has been widely analyzed, the focus on aggregate outcomes often makes it difficult to explore specific mechanisms through which the two are linked. Detailed household survey data allows us to examine changes in household landholdings and labor market behavior that are generated by reforms.

Our paper also employs an innovative empirical strategy. While several recent papers have exploited the random assignment of borders for institutional variation (Michalopoulos & Papaioannou 2011), sampling blocks that are linguistically similar but not immediately geographically adjacent allows us to address the concern raised by Bubb (2011) that there is little de facto variation in property rights across state borders, even if there is de jure variation.

This paper is organized as follows: Section 2.2 provides background on tenancy reform, a brief review of the literature on the economic impact of land reform, and a description of the natural experiment. Section 2.3 presents a theoretical framework used to generate predictions about tenancy reform. Section 2.4 introduces the data and discusses the empirical strategy. Section 2.5 provides the empirical results and Section 2.6 concludes.

2.2 Background

In this section we provide background on key points relevant to the analysis. First, we describe land relations in India at Independence and the subsequent tenancy reforms that we analyze. Next, we consider existing evidence on the effects of land reform. Finally, we describe the language-based state reorganization that we exploit in this analysis.

2.2.1 Land relations in India

The social and economic structure of village India is intrinsically tied to the caste system. Hindus, who make up over 80% of India's population, are born into a caste. Castes are endogamous groups defined by closed marriage and kinship circles.

Historically, the caste system also defined household occupation with land-ownership restricted among lower castes. Prior to British rule in India, inheritance determined land rights and land sales were extremely rare. While the British introduced new forms of land taxation, these changes did not disrupt the caste-land relationship, and rather worked to strengthen the correlation between landlessness and a low position in the caste hierarchy.

Two main forms of land taxation were introduced by British administrators. The first was the zamindari system, under which revenue liability for a given jurisdiction was assigned to a landlord who was empowered to collect revenue and enforce the payment of taxes. Zamindars were essentially awarded property rights for a village or group of villages (Banerjee & Iyer 2005). The second system was ryotwari, in which every registered landholder was recognized as a proprietor with the right to sell or transfer the land, and assured of permanent tenure as long as land revenue was paid. However, land taxes were high, and over time there were a significant number of both distress land sales and land appropriations by moneylenders when debt repayments were not made.

At Independence, India's large landowners were typically drawn from the upper castes. There were two main categories of tenants. First, there were occupancy tenants who enjoyed permanent heritable rights on land, security of tenure and could claim compensation from landlords for any improvement on the land. These were typically the middle and lower castes (often grouped as Other Backward Castes or OBCs). Second, tenants at will did not have security of tenure and could be evicted at the will of the landlord. This class consisted of the lowest castes and tribal households (grouped as Scheduled Castes and Tribes or SC/ST).

Quantitative and qualitative evidence collected in the early post-independence period emphasized that lower castes were largely landless laborers, servants, or tenants for the upper castes: e.g., in Tamil Nadu, 59% of the members of one upper caste were reported to be either landlords or rich peasants, while only 4% of the untouchable caste were landlords (Srinivas 1966, Sharma 1984). This translated into widespread landlessness – by 1956 estimates suggest that roughly one in every third rural household was landless, with the prevalence much higher among lower castes (Kumar 1962, Shah 2004).

Such statistics provided a significant impetus to land reform efforts in India post-Independence, to which we now turn.[2]

[2]The design of land reform in developing countries has long been a major preoccupation of policymakers and academics. A 1975 World Bank policy paper strongly supported redistributive land reform, with an emphasis on transferring land to more productive users and promoting owner-operated farms. However, implementation of these reforms has varied widely over the post-World War II period. While land reform was relatively successful in much of Asia where land relations were characterized by tenants cultivating landlord estates, in areas dominated by the hacienda system — in which tenants work on the landlord's farm and in turn receive a small plot of their own — it was largely stymied (Deininger & Binswanger 1999).

2.2.2 Land reform: Policy and existing evidence

In India, the constitution decreed that land policy was a state subject, and soon after Independence states began enacting such reforms, largely passed between 1950 and 1972. This wave of legislative activity included several major initiatives: the abolition of intermediaries, the imposition of land ceilings, and tenancy reforms.

The first type of reform, abolition reforms, abolished the zamindari system. Following the reforms, former tenants were now in a direct relationship with the state, rather than with a feudal lord, but this afforded relatively few immediate benefits. Even worse, abolition reforms often led to large-scale ejecting of "tenants-at-will, undertenants and sharecroppers." Since the laws abolishing zamindari allowed for retention of land for personal cultivation, many landholders responded by expelling tenants in order to increase this exempted area (Appu 1996).

The second class of reform included legislation that placed a ceiling on legal landholdings. Ceiling reforms, however, were typically weakened by provisions that set a high ceiling, established a large number of exceptions to the stated limit on landholdings, and offered no clear process by which to identify holders of surplus land or proceed against them (Rajan 1986, Radhakrishnan 1990).[3] Moreover, land that was redistributed was often in small plots and of poor quality, requiring substantial (and likely unaffordable) investments prior to cultivation (Herring 1991).

The final set of reforms – tenancy provisions that regulate relationships between tenants and landlords or, in some cases, render tenancy illegal – are widely identified as the best implemented reforms, characterized by less manipulation and fewer administrative bottlenecks (Eashvaraiah 1985, Herring 1991). However, even in this case, several authors note that larger tenants are the primary beneficiaries of tenancy provisions and differential eviction of informal tenants is common (Appu 1996).

The historical literature has elaborated extensively on the challenges encountered in implementing tenancy reform. Eashvaraiah (1985) in his analysis of Andhra Pradesh argues that the 1950 tenancy reform in effect created two classes of tenants, since those who were already evicted to avoid previous reforms were not reinstated and remained landless. Similarly, Pani (1983) argues that the implementation of land reform in Karnataka led to a large number of former tenants becoming agricultural laborers. Das (2000) contends that land reform resulted in tenants with substantial rights obtaining freehold occupation, while "in-

[3]Mearns (1999) argues that ceiling reforms achieved little because of the prevalence of loopholes and the bribing of record keepers or falsification of land records; see also Herring (1970) and Bandyopadhyay (1986).

ferior tillers," defined as inferior tenants, sharecroppers, contract farmers or paid laborers, lose access to cultivable land entirely.

Thus there are several reasons to focus on tenancy reform in this analysis. First, the previous literature generally suggests this was the only successful type of land reform, though certainly not without challenges. Second, this emphasis is consistent with the recent re-orientation of the broader land reform agenda towards a focus on the potential of land rental markets, appropriately regulated, as a means to provide the poor with some access to land (Deininger & Binswanger 1999).

Third, the design of tenancy laws implied that their impact would systematically vary with a household's initial tenurial security and access to credit. In almost every state, tenancy laws granted landowners rights of resumption for "personal cultivation", while tenants who remained on non-resumable tenanted land were eligible for ownership rights. In setting the land price, states either directly established a price or on occasion subsidized the market price. The design of the legislation thus ensures that the impact of land reform will be highly heterogeneous across pre-reform landownership status, which is closely linked to the historic caste structure.

We conclude with a review of quantitative studies on land reform in India. Banerjee, Gertler & Ghatak (2002) analyze Operation Barga, a program that encouraged tenancy registration in West Bengal and find that it lead to significant increases in agricultural productivity. However, Bardhan, Luca, Mookherjee & Pino (2011) find no clear evidence of reductions in inequality. A broader literature uses state-level variation in land reform to estimate its effect. Using cross-state evidence, Besley & Burgess (2000) find significant correlations between land reform and poverty reduction, while Conning & Robinson (2007) show that tenancy rates did fall as a result of land reform. Finally, Ghatak & Roy (2007) argue that land reform has no significant impact on land inequality as measured by the Gini coefficient.

There is also widespread evidence that, as we argue here, capital market imperfections play an important role in determining the structure of land markets and the impact of policy reforms on that structure. A basic empirical regularity indicative of the prevalence of these imperfections is the persistence of large land plots despite the well-documented negative land size-productivity relationship (Ray 1998).[4] Other evidence supporting this hypothesis includes the fact that the average land sale is a distress sale, thus creating a "market for

[4] This suggests that capital market frictions prevent landowners from extracting the full surplus from the sale of their land and thus inhibit sales that would otherwise be optimal.

lemons" problem that inhibits efficient sales (Rosenzweig & Binswanger 1993). Finally, land sale price often excludes the collateral value of land as buyers may have to mortgage land in order to purchase it (Binswanger, Deininger & Feder 1993, Deininger & Binswanger 1999). This leads potential buyers to undervalue the land, rendering the land market even thinner.

Finally, several recent studies examine the political economy of land reform. Mookherjee & Bardhan (2010) find evidence at the local level in West Bengal that the intensity of political competition (rather than party ideology) drives the incidence of land reform. Anderson, Francois & Kotwal (2011) present evidence that even post-land reform, landowners benefit from clientelist structures that they use to maintain political power and limit the implementation of policies that would redistribute income away from them. Similarly, for Mexico, de Janvry, Gonzalez-Navarro & Sadoulet (2011) argue that the left-wing party favored partial land reform over full land reform as it helped maintain a sufficiently large voter base of relatively poor voters. By documenting the pattern of gainers and losers, our study provides evidence that is useful in analyzing these political economy questions.

Against this background, we describe the institutional factors that we exploit in our analysis.

2.2.3 State reorganization and tenancy reform in South India

At the founding of India in 1947, its administrative structure reflected the history of expansion of the British East India Company and subsequently the British colonial government. Southern India was comprised of five states. Hyderabad and Mysore had been princely states under British rule, governed by local rulers with indirect colonial control via a British resident.[5] Travancore and Cochin were progressive princely states located on the southwest coast. The remainder of South India was directly ruled under the Madras presidency. The land tenure system in these states was a mix of Zamindari and Ryotwari, but the sub-district administrative unit of a block typically had a single land tenure system. Our unit of analysis is the block.

In the post-independence period, a movement grew to redraw state borders along linguistic lines. Based on the recommendations of a national commission, South India was divided into four linguistically unified states in 1956: Andhra Pradesh (AP), a largely Telugu-speaking state, was created from Hyderabad and the Telugu-majority areas of the Madras

[5]Hyderabad had originated as the territory of a Mughal governor who established control over part of the empire's territory in the Deccan plateau. Mysore emerged out of the defeat of the kingdom of Tipu Sultan in the early 19th century.

presidency. Karnataka (KA), intended to be predominantly Kannada-speaking, was created by the merger of Mysore and Kannada-speaking areas of Hyderabad and the Madras and Bombay presidencies. Kerala (KE), predominantly Mayalayam-speaking, encompassed the princely states of Travancore and Cochin and parts of the Madras presidency. Tamil-majority areas of the Madras presidency constituted the new state of Tamil Nadu (TN).

Districts were assigned to states primarily on the basis of the majority language spoken, but also in order to fairly assign valuable cities and ports, reasoning that was explained in great detail in the report produced by the commission (Government of India 1955). Figure 1 shows the borders of the new South Indian states overlaid on the previous state borders.

The state reorganization commission largely kept the sub-state administrative units of districts and blocks unchanged, identifying configurations of linguistically similar and geographically contiguous districts that would form a state. In some cases, however, blocks were reassigned across districts. Inevitably, on the borders of the new states, there were a number of cases in which two blocks with similar climate, geography and linguistic composition were separated into different states. Typically, these block pairs were previously part of the same state and possessed a shared political and administrative history.

Our identification strategy exploits the presence of such block pairs, under the assumption that shared history and linguistic (and caste) structure renders one block within the pair an appropriate control group for the other. On the latter point, it is relevant that in South India kinship structures and caste groups are defined within linguistic groups (Trautman 1981). Accordingly, blocks with similar linguistic comparison may plausibly be considered to have similar caste structures. Thus our sampling and survey strategy seeks to ensure that the two key features that determined pre-reform land distribution, caste structure in the village and political and administrative history, are held constant across the two blocks that are matched into a pair (on this, also see Section 2.4).

Next, data on tenancy reform in Southern India before and after the states' reorganization report is assembled from a variety of historical sources. Appendix Table 2.7 provides a summary of the number of tenancy reforms before (Pre) and after the states' reorganization (Post) in 1956 in the sampled districts, broken down by the number of each type of reform (abolition, ceiling and tenancy). Appendix Table **??** lists the dates and provisions of tenancy reforms by state. In general, tenancy reforms include measures that seek to enhance and codify tenants' rights to use their lands in specific ways; measures to prohibit eviction or the resumption of land use by the landlord; and in some states, legislation that grants full ownership rights to tenants. As we discuss in Section 2.4, a count measure of tenancy reforms will be employed as the primary independent variable.

The tables show that Kerala undertook the most land reform, and by the end of the period had prohibited tenancy. Andhra Pradesh and Tamil Nadu both experienced intermediate levels of land reform, and districts in Karnataka experienced the lowest cumulative levels of land reform. In all four states, provisions on maximum rent and tenants' rights to purchase land disincentivized tenancy arrangements (Appu 1996). In order for our identification strategy to generate accurate estimates of the impact of land reform on economic outcomes, we need to impose the assumption that other policies generating large shifts in rural landownership patterns do not meaningfully differ across states. Further discussion of the identification assumptions for the primary analysis can be found in Section 2.4.2.

2.3 Conceptual framework

Tenancy reforms can best be conceptualized as strengthening the rights of tenants. To capture the impact of this in theory, we develop a model in which landowners lack thte skill to farm land directly and thus choose whether to sell or rent their land. We consider the impact of a reform that allows tenants to capture a larger fraction of the surplus generated by land. While this makes tenants better off, landowners may choose to sell more land, thus altering patterns of land ownership, labor demand and wages.

2.3.1 Basics

There are three groups comprising a population: a measure π of landlords who owns all of the land and two groups of potential cultivators. The landlords own a measure $L < 1$ of land which we assume cannot be farmed directly, and land ownership is uniform among the landlord class. The technology matches one unit of land to one cultivator. We normalize the size of the group of cultivators to one.

The first group of cultivators, a fraction γ, have access to the capital market or some other form of wealth so that they can offer to buy land. In our data, this group will mainly comprise OBC households, but it could include some SC/ST households. The second group of cultivators, a fraction of $(1-\gamma)$, cannot buy land but can be taken on as tenants.

Whether as a tenant or an owner, a cultivator can employ labor on the land to generate output:

$$\theta \frac{1}{\eta} \ell^{\eta}$$

where $\eta < 1$ and $\theta \in [\underline{\theta}, \bar{\theta}]$ is an idiosyncratic productivity parameter which can be thought of as a cultivator's ability or access to relevant human capital. For simplicity, we assume that the distribution of ability is the same in each farmer group and denote this by $G(\theta)$.

Labor can be hired in a competitive labor market at a wage of w. It is supplied by cultivators who are neither tenants nor owners; there is always such a group since we have assumed that $L < 1$.

Let:

$$\pi(\theta, w) = \arg\max_{\ell} \left\{ \theta \frac{1}{\eta} \ell^{\eta} - w\ell \right\} = \frac{1}{\eta} \theta^{\frac{1}{1-\eta}} w^{-\frac{\eta}{1-\eta}}.$$

be the surplus generated by the land. Note that labor demand for a type θ cultivator is $(w/\theta)^{-\frac{1}{1-\eta}}$.

We will suppose that the same surplus is generated by either landlords or tenants and that the main issue is how institutions affect the distribution of this. In the event of selling the land, we suppose that the tenant can raise sufficient capital to pledge a fraction β of the surplus to the owner. Under tenancy, we suppose that the landlord can set the rent to earn a fraction α of the surplus. A key ratio affecting the analysis is α/β, i.e. the relative attractiveness of tenancy and selling. In an economy with highly imperfect capital markets and where the landlord has power over tenants, we would expect $\alpha/\beta > 1$.

2.3.2 Equilibrium

We are interested in two equilibrium decisions. First, the landlord decides how to divide his land between parcels to sell and parcels to rent out. Second, the labor market equilibrium generates the wage given this decision.

The landlord will decide how much land to rent out and how much to sell based on the ability of the farmer. Let

$$\hat{\theta}(x) = \left(\frac{\alpha}{\beta}\right)^{\frac{1}{1-\eta}} x \equiv \phi x$$

as the level of productivity that makes a landlord indifferent between selling and renting to a tenant of productivity level x.[6] If $\alpha/\beta > 1$, then $\hat{\theta}(x) > x$ which implies that the marginal cultivator who buys land will be more productive than the marginal tenant. So policies which encourage land sales will tend to drive up overall agricultural productivity.

[6]It is derived from

$$\beta\pi\left(\hat{\theta}(x), w\right) = \alpha\pi(x, w).$$

The landlord will sell some land and rent some land. Since he is assumed to be unable to directly farm any land, the least productive tenant who farms land, x, is defined from:

$$L = [1 - G(\phi x)]\gamma + (1-\gamma)[1 - G(x)]. \tag{2.1}$$

The first expression here is the land that is sold while the second is land that is rented. All the most productive cultivators farm land and the least productive are laborers. Note that:

$$\frac{\partial x}{\partial \phi} = -\frac{g(\phi x)\, x\gamma}{[\gamma g(\phi x)\,\phi + (1-\gamma)\, g(x)]} < 0. \tag{2.2}$$

Observe also using (2.2) that $\partial(\phi x)/\partial\phi > 0$. This says that the more that can extracted from tenants relative to sellers, the lower the productivity of the marginal tenant that is given land. The productivity gap between the marginal tenant and marginal owner cultivator also increases. This is because there is a switch towards tenants and away from selling the land.

The equilibrium wage solves:

$$\begin{aligned} 1 - L &= \gamma \int_{\phi x}^{\bar{\theta}} -\pi_w(\theta, w)\, dG(\theta) + (1-\gamma)\int_{x}^{\bar{\theta}} -\pi_w(\theta, w)\, dG(\theta) && (2.3)\\ &= w^{-\frac{1}{1-\eta}}\tilde{\theta}(\phi, x)\ , && (2.4)\end{aligned}$$

where $\tilde{\theta}(\phi, x) = \left[\gamma \int_{\phi x}^{\bar{\theta}} \theta^{\frac{1}{1-\eta}} dG(\theta) + (1-\gamma)\int_{x}^{\bar{\theta}} \theta^{\frac{1}{1-\eta}} dG(\theta)\right]$ be a measure of the average productivity of landlords and tenants. For future reference, observe that

$$\begin{aligned} \frac{d\tilde{\theta}(\phi, x)}{d\phi} &= \frac{\partial\tilde{\theta}(\phi, x)}{\partial\phi} + \frac{\partial\tilde{\theta}(\phi, x)}{\partial x}\cdot\frac{\partial x}{\partial\phi} && (2.5)\\ &= -\frac{g(\phi x)\, g(x)\,\gamma(1-\gamma)}{[\gamma g(\phi x)\,\phi + (1-\gamma)\, g(x)]} x^{\left(1+\frac{1}{1-\eta}\right)}[\phi - 1] \gtrless 0 \ \text{ as } \phi \gtrless 1. \end{aligned}$$

Whether average productivity rises or falls depends on whether the marginal tenant is more or less productive than the marginal owner of land.

An equilibrium in the land and labor market is a pair $(x^*(\phi), w^*(\phi))$ which solves (2.1) and (2.3). To explore the effects of tenancy reform, we are interested in how these depend on ϕ.

2.3.3 Tenancy reform

We now consider what happens when there is a reform that makes tenancy less attractive. We model this as a reduction in ϕ due to α having fallen. In other words, tenancy reform makes surplus extraction from tenants more difficult.

The model makes a number of predictions about the impact of this shift on landholding and wages, summarized as follows.

Model Predictions: *Suppose that tenancy reform reduces ϕ. The model predicts the following equilibrium responses:*

1. *An increase in landholding among the sub-group of the population with better capital market opportunities.*
2. *A reduction in tenancy.*
3. *An increase (decrease) in the agricultural wage if $\phi > (< 1)\, 1$.*

All of these effects of tenancy reform follow intuitively from the analysis above. By making tenancy less attractive, landlords sell more land to the group of cultivators who have the resources to purchase land.

The impact on wages is ambiguous a priori and depends on the initial conditions. In cases where the extractive power of landlords is strong then there will be a preference for tenancy even when the marginal tenant is fairly unproductive. In such cases wages will tend to rise with tenancy reform which reduces the power of landlords and encourages them to sell land which finds its way into the hands of relatively more productive farmers. This increases labor demand and hence wages. However, in cases where landlords are initially weak then the opposite would be the case.

The model can be used to explore the impact of tenancy reform on land inequality. A fraction

$$\beta_L(\phi) \equiv \frac{[(1-\gamma) + \gamma G(\phi x^*(\phi))]}{1+\pi}$$

are landless among whom $\frac{(1-\gamma)[1-G(x^*(\phi))]}{1+\pi}$ are tenants. A fraction $\frac{\pi+\gamma(1-G(\phi x^*(\phi)))}{1+\pi}$ of the population owns land. This can be decomposed into a fraction of owner-cultivators:

$$\beta_C(\phi) \equiv \frac{\gamma(1 - G(\phi x^*(\phi)))}{1+\pi}$$

which is decreasing in ϕ. The size of the landlord group remains fixed at π and, assuming that they sell land in equal numbers, their share of the land is:

$$\frac{[1-\gamma[1-G(\phi x^*(\phi))]]}{\pi}$$

which is increasing in ϕ.

Putting this together, it is straightforward to see that a reduction in ϕ leads to a new land distribution which Lorenz dominates the initial distribution. Hence, a wide variety of inequality measures, such as the Gini coefficient, should show a reduction in land inequality after tenancy reform.

To map the model further onto the data, note that we expect caste membership to map crudely onto our two cultivator sub-groups. Specifically, suppose that $\gamma = \gamma_{SC/ST} + \gamma_{OBC}$, then we would expect that $\gamma_{OBC} > \gamma_{SC/ST}$. While land ownership should rise in both groups, we expect this to be a larger effect for OBCs. Moreover, reductions in tenancy should be larger for the SC/ST group with a greater increase in participation as agricultural laborers. Land inequality between castes may increase as result of tenancy reform since OBC households will benefit disproportionately. Average income among the cultivator group J is:

$$\begin{aligned} \mu_J(\phi) &= w^*(\phi)\left[\gamma_J G(\phi x^*(\phi)) + (1-\gamma_J) G(x^*(\phi))\right] \\ &+\frac{\beta}{\eta}\left[w^*(\phi)\right]^{-\frac{\eta}{1-\eta}}\left[\gamma_J \int_{\phi x^*(\phi)}^{\bar{\theta}} \theta^{\frac{1}{1-\eta}} dG(\theta) + \phi(1-\gamma_J)\int_{x^*(\phi)}^{\bar{\theta}} \theta^{\frac{1}{1-\eta}} dG(\theta)\right]. \end{aligned}$$

The effect of a reduction in ϕ is ambiguous in sign for each group when groups differ in γ_J.

2.4 Data and empirical strategy

Our analysis makes use of multiple datasets. This section begins by describing the data and then outlines and justifies the empirical strategy employed.

2.4.1 Data

Tenancy reform data

Section 2.2.3 provided background on tenancy reform in the states of interest. A complete index of specific provisions enacted as part of tenancy reforms includes minimum terms of lease; the right of purchase of nonresumable lands; the right to mortgage land for credit;

mandatory recording of tenant names; limitations on the landlord's right of resumption; caps on rent; temporary protection against eviction or prohibition of eviction; prohibition of eviction for public trusts; the establishment of a system of processing land titles; the extension of formal tenancy to more classes of tenants; and the extension of full ownership rights to tenants.

Our primary definition of land reform follows Besley & Burgess (2000) and assumes that each piece of legislation represents a separate land reform event, and therefore is presumed to have an additional, cumulative impact on the distribution of land. We term this measure *Tenancy Index A*. The assumption underlying construction of this index may be violated if passage of additional legislation reflects simply the fact that earlier legislation was incomplete or ineffective, or if some states enact land reform incrementally while others enact only a few broad pieces of legislation.

To address this concern, we also report results for a second measure of tenancy reform denoted *Tenancy Index B*. This measure directly indexes the provisions enacted within the broad set enumerated above. Each district is assigned a dummy variable equal to one if the district experienced this type of reform, and the total score for tenancy is equal to the sum of these dummy variables.

Household and village survey

d Our sample includes nine boundary districts in the four Southern Indian states. Three sets of two adjacent districts constituted three separate pairs, and three adjacent districts (Kolar, Chittoor and Dharmapur) are compared pairwise, generating three additional pairs. Thus in total, there are six pairs of districts with four in the same princely state prior to 1956. Within each district pair, blocks were matched on linguistic similarity using a linguistic index based on 1991 census data on the proportion of the population speaking each one of the eighteen languages reported spoken in the region (for further details, see Appendix).

The language match index sought to identify block pairs separated by the post-1956 state boundaries where the difference across blocks in proportion population speaking each language is minimized. Within a district pair, the three independent (i.e., non-overlapping) pairs of blocks that were linguistic best matches were selected yielding 18 matched pairs of blocks (three pairs of blocks for each of six pairs of districts). The match quality indices for these block pairs are, on average, one and a half standard deviations lower (i.e., a closer match) than the mean.

The outcome variables were measured in a series of interlinked surveys conducted in the sampled villages in 2002. In each of a randomly selected 259 villages, twenty household surveys were conducted, yielding a total sample of 5180 households. Households were randomly selected, with the requirement that at least four households were SC/ST households. The survey collects data on familial structure, occupation, landholdings, and assets, as well as political knowledge and participation.

The second data set comprises data collected in 522 villages at a village-wide participatory rural appraisal (PRA) meeting at which attendees were asked to provide information about the caste and land structure in their villages, including the name of all castes represented and whether they were SC/ST, the number of households that belong to each caste, and the number of households falling into each one of a number of landowning categories. The same meeting was also used to obtain information from villagers about prevailing agricultural and construction wages. This methodology has previously been employed successfully in India to obtain data about public goods and recent public investments (Duflo, Chattopadhyay, Pande, Beaman & Topalova 2009).

The sampled villages are then linked to landholding data at the block and village level drawn from the 1951 census. The 1951 census reported the number of households in several land-owning/occupational categories (landlords, independent cultivators, tenants and landless laborers) by village, as well as data about literacy and the male and female population in the village. We are able to match 302 of the 522 villages in our sample.

2.4.2 Identification strategy

To examine the impact of tenancy reform we estimate two main specifications:

$$Y_{vp} = \beta_1 R_{vp} + \beta_2 X_{vp} + \gamma_p + \epsilon_{vp} \tag{2.6}$$

$$Y_{ivp} = \beta_1 R_{vp} + \beta_2 R_{vp} O_{ivp} + \beta_3 R_{vp} S_{ivp} + \beta_4 O_{ivp} + \beta_5 S_{ivp} + \beta_6 X_{vp} + \gamma_p + \epsilon_{ivp} \tag{2.7}$$

Y_{vp} denotes a inequality measure for village v in pair p and Y_{ivp} denotes an economic outcome for household i in village v and block-pair p. R_{vp} is an index of land reform for village v in block-pair p. O_{ivp} and S_{ivp} are indicators for the household being OBC or SC/ST, respectively, and X_{vp} denotes village-level controls. All regressions include a block-pair fixed effect γ_p.

For village-level regressions the standard errors would ideally be clustered at the level of the princely state and the state, comprising seven clusters. As inference employing clustered standard errors with a low number of clusters can be even more unreliable than inference using standard heteroskedasticity-robust standard errors, we estimate the specifications of interest without clustering and then re-estimate employing a wild bootstrap to bootstrap the T-statistics within each princely state-state cluster, following Cameron, Gelbach & Miller (2008).

The key identifying assumption is that, conditional on block-pair fixed effects, state assignment affects landowning and cultivation via tenancy reform. The second specification (2.7) requires a somewhat weaker identifying assumption, namely that state assignment affects landownership patterns across caste groups only via tenancy reform.

In Table 2.1 we present two checks on this identification strategy. First, we examine whether blocks that are matched according to linguistic closeness are more similar in pre-reform land structure.[7] To examine this, we employ the following procedure using the 1951 census data on village-level land structure: first, all possible matches between the sampled villages are created. Matches between villages in the same state are dropped, leaving only pairings across state lines. Some village pairs lie within the actual block pairs matched along linguistic lines, and some do not. To test whether the average difference in the percentage landless between villages in the matched block pairs is less than the average difference across all possible pairs of villages we estimate:

$$Dif_{j,k} = \beta Same_{j,k} + \mu_{s_j,s_k} \tag{2.8}$$

where $Same_{j,k}$ is an indicator variable equal to one when the villages are in a matched block-pair and zero otherwise, and μ_{s_j,s_k} is a dummy variable for matches between the states of village j and village k.

Column (1) Table 2.1 shows the results. On average, village pairs within matched blocked pairs are more similar than those not in matched pairs, with the difference in landless proportions about 11% less than the mean. Thus the matching process identifies block pairs that are more similar in both language and land structure.

Second, we examine whether assignment to different regimes of post-1956 land reform is uncorrelated with pre-period village characteristics within block-pairs: i.e., whether conditioning on linguistic similarity village assignment to states was quasi-random. Here, we estimate

$$R_{vp} = \beta x_{1951,vp} + \gamma_p + \epsilon_{vp} \tag{2.9}$$

where $x_{1951,vp}$ denotes covariates measured at the village level prior to the reorganization in

[7]In addition to land similarity we are also interested in caste structure similarity. However, due to the absence of micro-level caste data prior to the state reorganization, the test is restricted to land structure.

the 1951 census. R_{vp} denotes the number of tenancy reforms in village v of pair p post-1956 and, throughout, we include block-pair fixed effects, denoted γ_p.

The 1951 census provides data on the number of households in the village in specific livelihood classes, the number of literate men and women, and total population. Accordingly, (2.9) is estimated employing as independent variables the proportion of agricultural households that are tenants; the proportion of the population that is literate; and the proportion of the population that is engaged in agriculture.

The results are reported in Columns (2) through (7) of Table 2.1, employing both tenancy measures A and B as the independent variable. The wild bootstrap p-values are reported in brackets below the conventional standard errors. Columns (2) and (3) show no significant correlation between the 1951 literacy rate and post-1956 reform history. In Columns (4) and (5), there is some evidence of a negative correlation between the proportion of the population working in agriculture and subsequent land reform, significant using tenancy measure B.

Given the nature of land reform, whether this correlation holds for proportion of the agricultural population who are tenants is of most concern. However, columns (5) and (6) suggest that 1951 tenancy patterns in a village are not predictive of assignment to states with different subsequent land reform histories.[8] This suggests that the assignment of blocks to states was not intended to create a state more amenable to any particular land reform agenda, and within block pairs matched on the basis of linguistic and land structure, village assignment across states can be considered quasi-random. That said, throughout we report regressions with the 1951 demographic variable controls.

Finally, in all cases the pair fixed effects have significant explanatory power (we report the p-value for their joint significance in the tables), demonstrating that within-pair comparisons do help to control for unobserved heterogeneity across blocks.

2.5 Results

2.5.1 Land ownership by caste group

We start by using household data to examine the impact of land reform on differential land ownership by caste group. We estimate regressions of the form given by equation (2.7). The primary coefficients of interest are β_2 and β_3, capturing the heterogeneity of the effect of land reform across caste groups; upper caste households are the omitted base category. Standard

[8]Similarly, we observe no significant correlation for the other categories reported for the agricultural population (e.g., landlords, own-cultivators and landless laborers).

errors are clustered at the level of state-princely state-caste group, which is the level at which the interaction terms $R_{vp}O_{ivp}$ and $R_{vp}S_{ivp}$ vary. The outcome variables of interest are dummy variables for whether a household owns or leases land, and dummy variables capturing whether the primary source of income for the household is own-cultivation or agricultural labor.

Column (1) in Table 2.2 indicates that OBC households experience a significant increase in the probability that they own land as a result of tenancy reform, while SC/ST households show a significant decrease. Using the Panel A estimates at the mean of tenancy reform, the relative increase in the probability of OBC households owning land would be around 13 percentage points on a base probability of 60%. (The implied magnitude of the effect using Panel B estimates is nearly twice as large, though noisily estimated.) The relative decrease in the probability SC/ST households own land is around 20 percentage points on a base probability of 45%. Column (2) suggests no significant impact on the level of land leased, though the point estimate is, as predicted by theory, negative.

The coefficients on the dummy variables for the primary source of household income reported in columns (3) and (4) reinforce the finding of reduced tenancy for all social groups but differential impacts on land ownership. Column (3) shows that tenancy reform leads to relatively greater owner-cultivation among OBC households; using Panel A estimates, we observe an increase in probability of 9 points on a base probability of 31%. In contrast, owner-cultivation among upper caste and SC/ST households declines.

Column (4) shows that while OBC households are less likely to be dependent on agricultural labor after a tenancy reform, the probability that SC/ST households are dependent on agricultural labor increases by 15 percentage points on a base probability of 72% at the mean, for a proportional effect of 21% (Panel A estimates). There is a strong correlation between landlessness and dependence on agricultural labor as primary occupation; thus these coefficients capture the same underlying phenomenon of shifts in landlessness for SC/ST households, while employing different data.

In Panel B, which uses Tenancy Index B (calculated by indexing the number of separate provisions implemented) as the independent variable we observe a consistent pattern of coefficients. The absence of significant differences between the coefficients estimated in the two sets of regressions suggests that the observed pattern is not an artifact of the construction of tenancy reform variable.

These results reinforce the importance of examining the heterogeneous impact of tenancy reform at the household level, and suggest the effects plausibly depend on the extent to which potential cultivators can benefit from the possibility of becoming landowners as reform reduces the attractiveness of tenancy to landlords. An important part of measuring that is examining the impacts on agricultural productivity, to which we now turn.

2.5.2 Labor demand and wages

Our conceptual model predicts that tenancy reform may transfer land to more productive farmers, in which case overall labor demand and wages will increase especially where landlords initially have strong bargaining power. We now examine these two predictions.

Our first measure of labor demand is propensity of a household to engage in any paid agricultural labor. Households who do not report agricultural labor as their principal occupation may still provide agricultural labor if labor demand, and wages, are sufficiently high. In column (5) we see that tenancy reform increased participation in paid agricultural labor for all households, with the impacts largest for SC/ST households. (The impact on upper caste and OBC households is similar.) At the mean level of tenancy reform, the relative increase in the probability of SC/ST participation is around 17 percentage points on a base probability of 31%, a proportional effect of slightly over 50%. While larger in magnitude, this effect is consistent with the prior results and suggests that even households that did not report agricultural labor as their primary occupation were more likely to participate in the agricultural labor market post-reform.

In columns (6) and (7) we directly examine the impact on village agricultural wages, reporting specifications without and with trimming of the wage variable. The specification of interest is equation (2.6), and bootstrapped p-value are reported in brackets below the conventional standard error.

The results show that the censored measure of the daily agricultural wage increases by about 6% with each episode of land reform, or 42% at the mean level of land reform. An increase in the wage is consistent with the predictions of the model if $\phi > 1$, and also consistent with the results reported by Besley & Burgess (2000). In addition, the sizeable magnitude of the effect is in line with previous literature: Banerjee, Gertler & Ghatak (2002) estimate a positive effect of land reform on productivity of between 50% and 60%, implying an increase of comparable magnitude in the agricultural wage if the rural labor market is efficient.[9]

[9] As an additional robustness check, we estimate the coefficient on the interaction of tenancy reform and the pre-reform proportion of the population who are tenants, a proxy for the relative extractive power of landlords. Theory predicts that the wage effect should be larger where the extractive power of landlords is greater. The estimated interaction effect is positive, as predicted, and close to significant employing the wild

2.5.3 Overall land inequality

Next we examine whether, as predicted by the model, tenancy reform reduced overall land inequality. To do so, we make use of data on land distribution collected in participatory rural appraisal (PRA) meetings. These data are potentially noisier than household data but provide a valuable, supplementary account of shifts in overall land distribution.

In the PRA meeting, assembled villagers were asked to name for each caste the number of households that held no land, between 0 and 1 acres of land, 1 to 5 acres, 5 to 10 acres, 10 to 25 acres, or 25 or more acres. To calculate measures of inequality in landholdings, we assume that each household in a given category possessed the mean amount of land (e.g., a household holding between 1 and 5 acres is assumed to hold 3 acres).[10] The measures we examine include the Gini coefficient, the generalized entropy measure of inequality with α equal to 1, the ratios of total land held by percentiles 90/10 and percentiles 75/25, and the proportion of landless households.

Equation (2.6) is estimated again with both conventional heteroskedasticity-robust standard errors and wild bootstrap p-values reported in brackets. The results in Table 2.3 show that tenancy reform generally reduces overall inequality in land distribution, and the impact is substantial in magnitude.[11] First and possibly most important, there is a significant decline in the proportion of landless households that corresponds to a relative effect of around 10% at the mean level of tenancy reform. In addition, the Panel A estimates show a significant decline in the Gini coefficient of around 12%; a decrease in the Gini coefficient of this magnitude would move a village from the median level of inequality across all villages to the 25th percentile.

We observe even larger, though noisily estimated, reductions in the GE(1) measure of land inequality, around 20% at the mean level of tenancy in Panel A. A decline in the 90/10 ratio of around 20% at the mean level of tenancy is close to significant, and the decline in the 75/25 ratio is even larger (40%) and statistically significant. We observe no significant decline in between caste-group inequality.

Taken together, these results suggest an impact of tenancy reform which is consistent with the theoretical model laid out in the last section. There is a fall in overall inequality,

bootstrap p-value. Results are not reported but available by request.

[10] As our variables assume no dispersion within landholding categories they likely represent a lower bound on the true level of inequality. See Appendix for definitions of all measures.

[11] Estimating all the major results presented here with total reform as the independent variable results in coefficients of roughly equal magnitude, suggesting that abolition and ceiling reforms had no additional impact on village-level measures of inequality.

with land ownership increasing among OBC households and landlessness among SC/ST households. On the productivity front, we observe increases in wages and overall labor supply, with labor increases much higher for SC/ST households.

2.5.4 Robustness checks

Placebo tests A key challenge for the identification strategy is that tenancy reform may proxy for other state-level policies, and particularly for policies that differentially affect caste groups, benefiting middle castes at the expense of SC/ST households. Undeniably, the four states of interest did implement a variety of other different policies in this period. To provide some evidence about this variation, two regressions are estimated measuring the effect of assignment to a state with higher or lower levels of land reform on various measures of village- and household-level provision of public goods, and the interaction between land reform and caste dummies. For expositional ease, we only report results for Tenancy Index A.[12]

We start by examining village-level public good provision:

$$G_{vp} = \beta_1 R_{vp} + \beta_2 R_{vp} \times Pr_{vp} + \beta_3 Pr_{vp} + \gamma_p + \epsilon_{vp} \tag{2.10}$$

where G_{vp} is a dummy for whether the local government, denoted the gram panchayat or GP, provides a certain public good in the village and $R_{vp} \times Pr_{vp}$ is an interaction term with the proportion of SC/ST households in the village, denoted Pr_{vp}. Block-pair fixed effects γ_p are again employed, and standard errors are heteroskedasticity-robust.[13]

The results are shown in Columns (1) through (4) of Table 2.4. We observe no significant coefficient on either total reform or the interaction between reform and the proportion SC/ST, with the exception of a positive and marginally significant coefficient on the probability that the panchayat provides funds for repairs of the village school. This suggests that differential provision of public goods to villages with a higher or lower proportion of SC/ST households in states with more or less land reform is not a source of bias.

Next, we estimate the following equation at the household level:

$$G_{ivp} = \beta_1 R_{vp} + \beta_2 R_{ip} \times O_{ivp} + \beta_3 R_{vp} \times S_{ivp} + \beta_4 O_{ivp} + \beta_5 S_{ivp} + \gamma_p + \epsilon_{ivp} \tag{2.11}$$

[12]The results are similar for Tenancy Reform B.

[13]These standard errors should also be estimated using a wild bootstrap. However, the objective in this test is to test for the presence of a null effect; accordingly, using standard errors that are biased toward zero is a more stringent test of this assumption.

where G_{ivp} is a dummy for the provision of governmental assistance to that household or the colony in which the household resides. Analogous to the main specifications, this equation is estimated with block-pair fixed effects and interaction terms for caste group. Standard errors are clustered at the state-princely state-caste group level.

The results are shown in Columns (5) through (7) , using as the dependent variable a dummy for whether the household received government aid for construction or electricity, whether the colony received infrastructure provided by the government, and whether the household is eligible for a BPL card. The results show a coefficient on the interaction between SC/ST and total reform that is positive and sometimes significant: in other words, SC/ST households are more likely to receive government assistance in states that have more land reform.

These results indicate that insofar as differential provision of public goods or governmental assistance in states with more or less reform introduces bias to our results, the bias seems to be towards finding a positive effect on the welfare for SC/ST households. This could also be interpreted as a corollary of the increased landlessness for SC/ST households rendering them eligible for such assistance.

Alternative specifications A final set of robustness checks re-estimates the primary equation of interest (2.7) employing an index of total land reform, rather than tenancy, as the independent variable. The objective of this regression is to evaluate whether the observed pattern of effects for tenancy reform is also evident for overall land reform.

The results are shown in Table 2.5, and the coefficients are entirely consistent with the previous results. In fact, there are no significant differences between the coefficients estimated using tenancy reform and total reform. This suggests that as concluded by the previous qualitative literature, tenancy reforms are the only measures that are effective in altering land ownership patterns. In fact, the estimated impacts of tenancy legislation and all land reform legislation are statistically indistinguishable.

2.6 Conclusion

Poor rural economies are second-best in many ways. It is no surprise, therefore, that tracing the impact of a single dimension of reform can be complex. The analysis in this paper has exploited a natural experiment due to the 1956 state reorganizations in India to evaluate the impact of tenancy reform at the village and household level over a long time horizon.

While tenancy reforms were implemented with the goal of strengthening the position of tenants, several equilibrium responses need to be considered. In this context, the reforms did produce significant and highly persistent shifts in land distribution. However, the benefits were lopsided and favored relatively wealthy tenants, while SC/ST households saw a decrease in land holdings and generally became more reliant on agricultural labor.

On the other hand, there is evidence of a large increase in agricultural wages due to an increase in demand for hired labor. This phenomenon could be due either to large landholders no longer relying on tenancy and/or a shift in the labor supply curve. Thus while the welfare impacts of tenancy reforms were substantial and long-lasting, their impact was heterogeneous between types of cultivators. These results can best be understood through the lens of a fairly standard model where owners of land are seeking the best opportunities for exploiting their land and there is a reduction in landlords' ability to extract surplus from tenants due to the reform.

The question of how best to regulate the land market is still a pressing one in many developing economies. Mexico has embarked on major experiments in rural land titling over the last decade (de Janvry, Gonzalez-Navarro & Sadoulet 2011). Rural land rights remain extremely limited in China, where the role of property rights in rural development is hotly contested and has become an increasing source of political unrest. In addition, many other developing countries face challenges in how to appropriately negotiate compensation for rural landowners when industrialization requires the purchase or expropriation of land (Bardhan 2011). In all such cases, it is essential to understand in detail, as we have done here, the equilibrium responses to reform and the way that these responses create gainers and losers. This can only be done employing a sufficiently long time horizon over which the full effects of reform become visible.

In a broad sense, our findings offer a stark reminder of the hazards of piecemeal policy reform in a second-best world. If tenancy persists in part due to a lack of credit market opportunities to become an owner-cultivator, then increasing the power of tenants may result in some tenants being forced to become landless laborers; the ultimate welfare impact for these tenants will depend on the strength of factor market shifts in equilibrium, primarily the wage response. The complexity of these general equilibrium effects should contribute to a recognition by policymakers that, while short-run political imperatives may provide the impetus for reform, the long-run economic changes are what matter for development.

2.7 Figures and tables

Figure 2-1: Map of sample districts

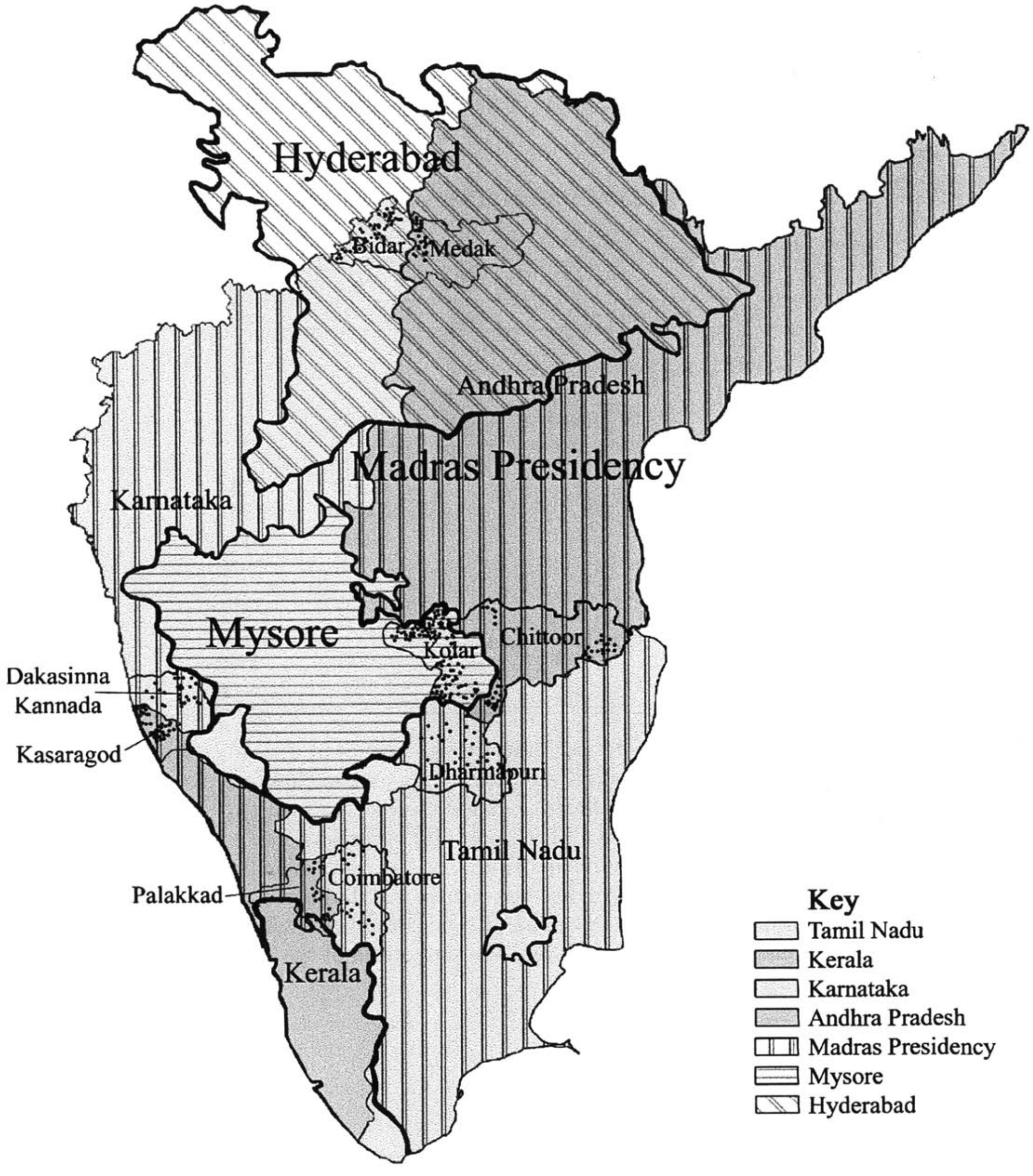


Table 2.1: Quasi-random assignment of villages to states

	All block pairs	Index of post-1956 tenancy reform					
		A	B	A	B	A	B
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Same pair	-.027						
	(.010)***						
Prop. literate		2.328	.883				
		(1.369)	(.581)				
		[.933]	[.724]				
Prop. agricultural				-3.267	-1.247		
				(.655)	(.287)		
				[.170]	[.000]***		
Prop. tenants						-3.562	-1.513
						(1.130)	(.488)
						[.733]	[.487]
Joint p-value pair FE		.000	.000	.000	.000	.000	.000
Obs.	10308	288	288	272	272	284	284

Notes: standard errors are in parentheses; asterisks indicate significance at 1, 5 and 10 percent levels. The first column tests whether block-pairs matched on a measure of linguistic closeness also have similar land structures; the regression included state fixed effects, and standard errors with two-way clustering at the level of the block and the paired block. Columns (2) through (4) regress an index of tenancy reform post-independence on demographic variables reported in the 1951 census, including block-pair fixed effects; heteroskedasticity-robust standard errors are reported as well as wild-bootstrap p-values reported in brackets. The independent variables are the proportion of the agricultural population that is a tenant, the proportion of the overall population that is literate and the proportion of the overall population engaged in agriculture.

Table 2.2: Impact of land reform on land ownership

	Land dummy (1)	Leased dummy (2)	Own cult. (3)	Agri. labor (4)	Agri. labor (ind.) (5)	Wage (6)	Wage trim (7)
Panel A: Tenancy Index A							
Tenancy reform	-.008	-.0006	-.016	.010	.008	4.050	2.711
	(.006)	(.004)	(.006)**	(.006)	(.002)***	(.414)	(.028)
						[.075]*	[.050]**
SC/ST x Tenancy	-.028	-.002	.006	.012	.017		
	(.015)*	(.002)	(.008)	(.006)*	(.002)***		
OBC x Tenancy	.019	-.003	.029	-.023	.001		
	(.009)**	(.002)	(.008)***	(.009)**	(.002)		
SC/ST	-.060	.007	-.348	.372	.112		
	(.077)	(.007)	(.041)***	(.033)***	(.013)***		
OBC	-.227	-.012	-.363	.373	.074		
	(.091)**	(.025)	(.076)***	(.086)***	(.031)**		
Joint p-value pair FE	0	0	0	4.720e-57	1.22e-163	0	0
Panel B: Tenancy Index B							
Tenancy reform	-.023	.003	-.035	.028	.015	5.211	3.424
	(.011)**	(.007)	(.009)***	(.010)***	(.004)***	(.935)	(.060)
						[.124]	[.129]
SC/ST x Tenancy	-.068	-.005	.002	.021	.028		
	(.021)***	(.004)	(.014)	(.010)**	(.004)***		
OBC x Tenancy	.035	.013	.063	-.070	-.012		
	(.037)	(.012)	(.031)**	(.037)*	(.007)***		
SC/ST	.155	.023	-.332	.325	.048		
	(.094)*	(.016)	(.064)***	(.048)***	(.034)		
OBC	-.310	-.126	-.556	.649	.164		
	(.257)	(.078)	(.219)**	(.254)**	(.029)*		
Joint p-value pair FE	0	0	0	3.55e-11	0	5.741e-40	0
Mean	.607	.097	.377	.438	.166	60.563	55.987
Obs.	2822	1940	2822	2822	15144	2867	2867

Notes: standard errors are clustered at the state-princely state-caste group level and reported in parentheses; asterisks indicate significance at 1, 5 and 10 percent levels. Wild bootstrap p-values are reported in brackets for outcomes measured at the village level. All regressions include block pair fixed effects. The dependent variables in Columns (1) through (4) are reported at the household level: a dummy for owning land, a dummy for leasing land, a dummy for being primarily dependent on own cultivation, and a dummy for being primarily dependent on agricultural labor. A large number of households gave no response to the question on leasing, leading to a large number of missing variables in that regression. Column (5) is an individual-level dummy denoting participation in non-agricultural labor, and Column (6) and (7) report the wage. Pre-reform controls included are the proportion of the agricultural population that are tenants, the proportion of the total population that is literate, and the proportion of the total population engaged in agriculture.

Table 2.3: Impact of land reform on inequality in land distribution

	Prop. landless (1)	Gini (2)	GE(1) (3)	90/10 (4)	75/25 (5)	BC(1) (6)
			Panel A: Tenancy Index A			
Tenancy reform	-.005	-.009	-.024	-.686	-.661	-.012
	(.006)	(.004)	(.009)	(.953)	(.400)	(.006)
	[.015]**	[.020]**	[.751]	[.203]	[.020]**	[.761]
Joint p-value pair FE	2.130e-07	4.107e-08	7.022e-07	.000	.006	3.722e-18
			Panel B: Tenancy Index B			
Alternate measure	-.002	-.012	-.030	-1.173	-1.439	-.009
	(.010)	(.007)	(.015)	(1.719)	(.688)	(.010)
	[.095]*	[.040]**	[.532]	[.174]	[.219]	[.592]
Joint p-value pair FE	2.565e-06	3.840e-07	2.164e-07	.000	.015	2.463e-18
Mean	.336	.527	.620	34.493	13.048	.211
Obs.	302	302	302	302	302	302

Notes: standard errors are in parentheses; asterisks indicate significance at 1, 5 and 10 percent levels employing conventional standard errors. Wild bootstrap p-values are reported in brackets. All regressions include block-pair fixed effects and controls for the pre-reform measures of the proportion of the agricultural population that are tenants, the proportion of the total population that is literate, and the proportion of the total population engaged in agriculture. Outcome variables are the Gini coefficient, GE(1) coefficient, 90-10 ratio, 75-25 ratio, and between-caste GE(1) ratio in land inequality, and the proportion of the population that is landless.

Table 2.4: Placebo tests

	School Repair	Center Repair	Health Salaries	Health Materials	Hh infra.	Colony infra.	BPL card
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Tenancy	-.010	.008	-.007	-.005			
	(.016)	(.012)	(.005)	(.005)			
Tenancy x SC/ST prop.	.030	-.031	.006	.006			
	(.041)	(.023)	(.005)	(.006)			
Tenancy					-.011	.043	.007
					(.002)***	(.006)***	(.010)
Tenancy x SC/ST					.011	.009	.022
					(.003)***	(.006)*	(.009)**
Tenancy x OBC					.007	-.027	.015
					(.005)	(.025)	(.016)
SC/ST					.095	-.052	.096
					(.013)***	(.025)**	(.076)
OBC					-.042	.223	-.126
					(.050)	(.169)	(.150)
Obs.	302	302	302	302	2822	2426	2822

Notes: standard errors are in parentheses; regressions including clustering are at the state-princely state-caste group level. All regressions include block-pair fixed effects. Asterisks indicate significance at 1, 5 and 10 percent levels. In Columns (1)-(4), the dependent variables are dummies for whether the panchayat provided any funds toward the specified educational or health public good, and SC/ST prop. int. is an interaction between the proportion of the village population that is SC/ST and the tenancy variable. Standard errors are heteroskedasticity-robust. In Columns (5)-(7), the dependent variables are dummies for whether a household received assistance in improving their home from a public assistance scheme; whether the colony in which the household lives received such assistance; and whether the household is eligible for a BPL card.

Table 2.5: Impact of total land reform

	Land dummy (1)	Leased dummy (2)	Own cult. (3)	Agri. labor (4)	Agri. labor (ind.) (5)	Wage (6)	Wage trim (7)
Total reform	-.009	-.001	-.019	.012	.009	4.032	2.665
	(.005)*	(.004)	(.006)***	(.006)**	(.002)***	(.432)	(.029)
						[.283]	[.269]
SC/ST x Total reform	-.037	-.0004	.005	.014	.017		
	(.015)**	(.003)	(.008)	(.006)**	(.002)***		
OBC x Total reform	.016	-.007	.031	-.024	.005		
	(.015)	(.004)	(.013)**	(.013)*	(.003)		
SC/ST	.099	.003	-.360	.322	.060		
	(.105)	(.017)	(.059)***	(.047)***	(.010)***		
OBC	-.255	.037	-.470	.454	.023		
	(.179)	(.047)	(.149)***	(.159)***	(.043)		
Joint p-value pair FE	0	0	0	5.51e-142	2.14e-137	0	0
Mean	.607	.097	.377	.438	.166	60.563	55.987
Obs.	2822	1940	2822	2822	15144	2867	2867

Notes: standard errors are clustered at the princely state-state-caste group level and reported in parentheses; asterisks indicate significance at 1, 5 and 10 percent levels. Wild bootstrap p-values are reported in brackets for outcomes measured at the village level. All regressions include block-pair fixed effects. The dependent variables in Columns (1) through (4) are reported at the household level: a dummy for owning land, a dummy for leasing land, a dummy for being primarily dependent on own cultivation, and a dummy for being primarily dependent on agricultural labor. A large number of households gave no response to the question on leasing, leading to a large number of missing variables in that regression. Column (5) is an individual-level dummy denoting participation in non-agricultural labor, and Column (6) and (7) report the wage. Pre-reform controls included are the proportion of the agricultural population that are tenants, the proportion of the total population that is literate, and the proportion of the total population engaged in agriculture.

2.A Appendix: Sampling methods

We selected four pairs of districts formerly in the same princely state that were incorporated into two different states. Bidar and Medak in Hyderabad were incorporated into Karnataka and Andhra Pradesh, respectively. In the Madras presidency, there are three such pairs: South Kanara (Karnataka) and Kasaragod (Kerala), Pallakad (Kerala) and Coimbatore (Tamil Nadu), and Dharmapuri (Tamil Nadu) and Chittoor (Andhra Pradesh).

Given that Mysore was completely incorporated into Karnataka, there are no district-pairs in which both districts were formerly part of Mysore state. However, Kolar district in Mysore / Karnataka was also surveyed, and matched on the basis of language, as detailed below, with Chittoor district in Andhra Pradesh and Dharmapuri in Tamil Nadu. All three districts form a contiguous geographic region, and they are matched pair-wise to generate three additional district pairs.

In order to select the block pairs employed in this analysis, blocks within the paired districts were matched on the basis of linguistic compatibility. For each block pair of block i and block j, a measure of linguistic compatibility $L_i(v_i, v_j)$ was constructed using the following formula. P_{li} denotes the proportion of the population in block i speaking a given language,[14] and N_i denotes the population in a given block. Thus L_i equals the sum of the difference in the proportion of population speaking each language across the two blocks, each weighted by the proportion of the population that speaks that language in both blocks taken as a whole. The minimum possible value of the index of linguistic compatibility, indicating the best possible match, is zero; the maximum is one.

$$L_i(v_i, v_j) = \sum_{l=1}^{18}(P_{li} - P_{lj}) * \frac{P_{li} * N_i + P_{lj} * N_j}{N_i + N_j} \tag{2.12}$$

For each district pair, the set of all possible block pairs is ranked and the top three unique pairs are chosen. Table 2.6 shows summary statistics for the quality of match for all possible block pairs for each pair of districts. On average, block pairs show the highest degree of linguistic compatibility across Kolar and Chittoor districts, and the lowest degree of compatibility in Coimbatore and Palakkad districts. The other four district pairs have similar levels of language matching. The high quality of the matches between Kolar and Chittoor and Kolar and Dharmapuri districts indicates that despite the fact that these district pairs

[14]The languages reported are Assamese, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Marathi, Mayalayam, Manipuri, Nepali, Oriya, Punjabi, Sanskrit, Sindhi, Tamil, Telugu, and Urdu.

were not previously part of the same princely state, their ethnolinguistic composition is comparable.

Table 2.6: Linguistic compatibility across district-pairs

District pair	Mean L_i	Median L_i	Std. dev.
Bidar-Medak	0.47	0.46	0.09
Chittoor-Dharmapuri	0.58	0.65	0.20
Dakasinna-Kasaragod	0.47	0.43	0.21
Coimbatore-Palakkad	0.74	0.73	0.13
Chittoor-Kolar	0.28	0.27	0.16
Dharmapuri-Kolar	0.52	0.57	0.19

Blocks are divided into village government units or gram panchayats (GPs), consisting of one to six villages. In the states of Andhra Pradesh, Tamil Nadu, and Karnataka, six gram panchayats were randomly sampled from each block selected. Gram panchayats in Kerala are larger than those in other states, and thus three GPs were sampled in each block in Kerala. All villages in each GP were sampled in AP, TN and KA if the GP had three or fewer villages; if there were more than three villages, then the village that was the home of the president of the gram panchayat was sampled in addition to two other randomly selected villages. (For the purposes of the sampling frame, villages with a population of less than 200 were excluded; all hamlets with a population over 200 are considered independent villages.) In Kerala, villages are again much larger and thus wards, the subunit of villages, were directly sampled. Six wards in each GP were randomly selected. This generates a total sample of 527 villages.

2.B Appendix: Inequality measures

The Gini measure is defined as follows, where l_i denotes the land owned by household i, r_i is the ranking of household i according to land holdings among all households in the village, $\bar{l}$ is mean land held in a village and n is the total number of households:

$$Gini = 1 + \frac{1}{n} - \frac{2}{\bar{l}n^2}\sum_{i=1}^{n}(n - r_i + 1)(l_i) \tag{2.13}$$

The general entropy measures with a=1 and a=2 are calculated using the following equations:

$$GE(a) = \frac{1}{a(a-1)} \left[\left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{l_i}{\bar{l}} \right)^a \right] - 1 \right] \tag{2.14}$$

2.C Appendix: Land reform in Southern India

Table 2.7: Summary statistics on land reform

State	District	Total reform Pre	Total reform Post	Abolition Pre	Abolition Post	Ceiling Pre	Ceiling Post	Tenancy Pre	Tenancy Post
KA	Bidar	6	3	3	1	0	2	3	2
AP	Medak	6	6	3	1	0	2	3	3
AP	Chittoor	5	6	0	1	0	2	5	3
TN	Dharmapuri	5	7	0	1	0	2	5	4
KA	Dakasina Kannada	5	3	0	1	0	2	5	2
KE	Kasaragod	5	10	0	2	0	1	5	9
TN	Coimbatore	5	7	0	1	0	2	5	4
KE	Palakkad	5	10	0	2	0	1	5	9
KA	Kolar	3	3	2	1	3	2	1	2

Note: the total number of reforms for Karnataka and Kerala, all post-1956, differs from the sum of the categories given that they incorporate legislation that can be jointly categorized. For Karnataka, the 1961 and 1974 acts include both tenancy reforms and land ceilings. For Kerala, the 1969 Kerala Land Reforms Act includes all three types of provisions.

Table 2.8: Land reform: Hyderabad

Year	Title	Description	Type
1950	Telegana Agency Tenancy and Agricultural Lands Act	Tenants received protected tenancy status; tenants to have minimum terms of lease; right of purchase of nonresumable lands; transfer of ownership to protected tenants in respect of nonresumable lands; as a result 13,611 protected tenants declared owners.[15] Also gave tenants ability to mortgage rented land for credit.[16]	Tenancy
1954	Amendment of Telegana Agency Tenancy and Agricultural Lands Act	Limits a landlord's right of resumption.[17]	Tenancy
1956	Tenancy Act (amended 1974)	Tenancy continues up to 2/3 of ceiling area; law does not provide for conferment of ownership rights on tenants except through right to purchase; confers continuous right of resumption on landowners.[18]	Tenancy

Table 2.9: Land reform: Madras Presidency

Year	Title	Description	Type
1929	Malabar Tenancy Act	Confers a qualified fixity of tenure on cultivation and a right to demand a renewal of lease. Also prescribed rates of "fair" rent. Since this act only took effect in the Malabar region of Madras Presidency, in our sample it only applies to Palakkad district.[19]	Tenancy
1954	The Malabar Tenancy Amendment Act	Prohibits eviction of tenants who have had land possession for 6 years; lowered the amount of maximum rent that could be paid.[20]	Tenancy
1955	The Madras Cultivating Tenants Protection Act	Prohibits any cultivating tenant from being evicted, except in the case of non-payment, but allows for resumption of up to one-half land if land leased out to tenant.[21]	Tenancy
1956	The Madras Cultivating Tenants (Payment of Fair Rent) Act	Abolishes usury and rack-renting;[22] Fixes the percentage of produce that can be charged as rent.[23]	Tenancy

Table 2.10: Land reform: Mysore

Year	Title	Description	Type
1952	Mysore Tenancy Act (Mysore Act XIII of 1952)	Restricted rent to 1/3 of crop; granted permanent tenancy rights to those who had occupied the land for 12 years or more. Also provided for the eviction of tenants for non-payment of rent and for resumption for self cultivation by landlord.[24]	Tenancy

Table 2.11: Land reform: Karnataka

Year	Title	Description	Type
1961	Land Reforms Act Amended 32 times (1965-2001)	Provides fixed tenure subject to landlord's right to resume one-half leased area; grants tenants optional right to purchase land on payment of 15-20 times the net rent; imposition of ceiling on land holders.[25]	Tenancy, Ceiling
1974	The Mysore Land Reforms Amendment Act	Imposition of ceiling on landholdings of 4.05-21.85 hectares (after 1972); removal of all but one of the exemptions from tenancy regulations;[26] reduces the landlord's right of resumption.[27]	Tenancy, Ceiling

Table 2.12: Land reform: Andhra Pradesh

Year	Title	Description	Type
1957	The Andhra Tenancy Act	A stop-gap measure to stay evictions of tenants in the Andhra area until new state-wide legislation could be drafted.[28] In our sample this act, and its amendment (listed below), only applies to Chittoor.	Tenancy
1971	Andhra Pradesh Record of Rights in Land Act	Provides for the recording of names of all occupants and tenants.[29]	Tenancy
1974	Amendment of Tenancy Act	Applied the 1956 tenancy laws to the whole state; reduced the maximum rent tenants paid; limits a landlord's right of resumption.[30] (In our sample this amendment only applies to Chittoor.)	Tenancy

Table 2.13: Land reform: Kerala

1957	Kerala Stay of Eviction Act	Provides temporary protection to tenants, kudikidappukars and persons cultivating land on minor sub tenures.[31]	Tenancy
1963	Kerala Land Reforms Act	Concedes tenants right to purchase land from landowners.[32] Amended 9 times (1969-1989)	Tenancy
1963	Kerala Tenants and Kudikidappukars Protection Act	Provides temporary protection to tenants in the matter of eviction.[33] and recovering of arrears of rent.	Tenancy
1966	The Kerala Prevention of Eviction Act (Kerala Act 12 of 1966)	Protected tenants against eviction; stopped recovery of rent arrears[34] from before April 1966.	Tenancy
1968	The Kerala Records of Rights Acts	Establishes records of land/tenancy rights.[35]	Tenancy
1969	The Kerala Land Reforms Amendment Act (Kerala Act 35 of 1969)	Conferment of full ownership rights on tenants; 2.5 million tenants could become land owners; right of resumption expires; imposition of ceiling on land holdings of 6.07-15.18 hectares (1960-1972) and of 4.86-6.07 hectares (after 1972); abolition of intermediary rights.[36]	Tenancy, Abolition, Ceiling
1972	The Kerala Land Reforms Amendment Act (Kerala Act 17 of 1972)	Changes the way the government processed land-titles; requires that statements be filed by large land holders.[37]	Tenancy
1976	The Kanam Tenancy Abolition Act (Kerala Act 16 of 1976)	Abolishes a form of intermediary.[38]	Tenancy
1989	The Kerala Land Reforms Amendment Act	Extends the benefits of tenancy and security of tenure to two more classes of tenants.	Tenancy

Table 2.14: Land reform: Tamil Nadu

1961	Madras Public Trusts Regulation of Administration of Agricultural Lands Act	Provides that no public trust can evict its cultivating tenants.[39] Limits the amount of land a public trust can personally cultivate.[40]	Tenancy
1969	Agricultural Land-Records of Tenancy Right Act	Provides for preparation and maintenance of complete record of tenancy rights.[41]	Tenancy
1971	Occupants of Kudiyiruppu Act	Provides for acquisition and conferment of ownership right on agriculturists, agricultural laborers, and rural artisans.[42]	Tenancy
1995	Amendment to the Tamil Nadu Cultivating Tenants Protection Act	Provides former cultivating tenants who had possession of land on Dec 1, 1953 the right to resume that land on the same term as held in 1953.[43]	Tenancy

Chapter 3

Sibling Rivalry: Ability and Intrahousehold Allocation in Gansu Province, China

3.1 Introduction

For decades, social scientists have analyzed the decisions households make about human capital accumulation and the implications of these decisions for individual economic outcomes. Given that the majority of education occurs in childhood, it is particularly important to understand the choices that parents make about education on behalf of their children. In multichild families, this entails not only identifying resources for education in the form of money or parental time, but allocating those resources among multiple children. The process by which these decisions are made remains poorly understood.

This paper provides evidence about the parental allocation of resources for education among children of varying endowments in a low-resource setting in rural China. The research question of interest is whether parents employ a compensatory or a reinforcing strategy in responding to innate variations in health and robustness among their children. However, direct estimation of the relationship between parental behavior and relative health or ability poses serious challenges. Any measurement of child characteristics will inevitably include a component of endogenous parental nurturing, and thus a parental preference may manifest itself in both better health for a given child and in overtly preferential treatment of the same child. This generates spurious evidence of a positive relationship between a child's endowment and parental investments.

The principal methodological contribution of this paper is to address the endogeneity of children's measured endowment by employing as an instrument a measure of exogenous variation in resource availability correlated with physical health. The instrument used is rainfall and grain yield in infancy for each child, an index of nutritional availability during a critical period of childhood development that substantially determines physical endowment. There is a broad consensus in the existing medical literature that malnutrition in the first years of life, particularly during the prenatal period and between birth and age three, has a substantial negative impact on physical and cognitive development (Pollitt, Gorman, Engle, Rivera & Martorell 1999, Grantham-McGregor & Ani 2001). Shocks to a child's nutritional intake in this period are correlated with endowment, but exogenous to other intrahousehold decision-making processes given the use of household fixed effects that absorb shocks to the household's overall budget constraint.

The results show a clear pattern of spending allocations favoring the child with lower endowment, consistent with a parental preference for equality that seeks to compensate for variation in endowment induced by early childhood environmental shocks. This pattern of preferential allocations holds across multiple measures of expenditure, and is robust to the inclusion of gender and sibling parity. In this context, educational spending seems to be employed as a tool to neutralize differences in endowment that would a priori be correlated with differences in expected income between children.

Previous literature examining intrahousehold allocation of resources to offspring has largely focused on the question of differential allocation to male versus female children, with a substantial literature establishing a pattern of preferential allocations to male children in both South and East Asia (Hazarika 2000, Behrman & Deolalikar 1990, Ono 2004). Other studies have examined the impact of the sex ratio of siblings on a child's education, finding that a child with more sisters has better health and education outcomes than one with more brothers (Garg & Morduch 1998, Morduch 2000), though the inverse relationship appears to hold in the United States (Butcher & Case 1994). A separate literature has focused on the relationship between birth order and the intrafamily distribution of resources (Parish & Willis 1993, Tenikue & Verheyden 2007, Bommiere & Lambert 2004).

A smaller literature has analyzed whether or not parents have a general preference for equality among their children. An early paper by Griliches presented evidence that parents attempted to limit intrafamily equality and attenuate preexisting differences in endowment, noting that the effect of IQ on schooling is significantly lower within families (Griliches 1979). Behrman, Pollak & Taubman (1982) examine familial allocations using twin data from the

U.S. and reject the pure investment model in which parents care only about the total return to educational expenditure, employing functional form assumptions on the parental welfare function. Using developing country data, Rosenzweig & Wolpin (1988) find that parents in Colombia attempt to compensate for the disadvantages suffered by children with lower weight at birth by a longer interval prior to the birth of the next child, though there is contravening evidence that healthier children are more often breastfed. Behrman (1988) finds evidence in India of a pro-male bias as well as parental inequality aversion, though such aversion declines in the lean season.

Two more recent papers analyzing the response of parental human capital investments to children's variation in endowments found that parents exhibit reinforcing behavior. Rosenzweig & Zhang (2009) find that parents exhibit higher educational expenditure for children of higher birth weight in China. They do not address the potential endogeneity of birth weight for siblings born as singletons, but find parallel results for twin pairs for which endogenous determination of birth weight can be ruled out. It is possible that these results cannot be easily extrapolated to non-twin siblings given that twins presumably compete more directly for parental resources and time than non-twins. Conti, Heckman, Yi & Zhang (2010) also employ data from China to estimate a structural model of parental resource allocation given multidimensional child endowments and find evidence of compensating investment in health but reinforcing investment in education. Their identification strategy relies on the assumption that early health shocks, defined as a reported episode of serious disease, occur randomly within twin pairs.

This paper makes several new contributions to the existing literature on parental intrahousehold allocation. It is the first study to estimate the response of parental allocations to quasi-exogenous variation in endowment without relying on the use of twin pairs, and the first to employ climatic shocks in infancy in an identification strategy to address the potential endogeneity of a child's endowment. The paper proceeds as follows. Section 2 outlines the theoretical framework for analyzing intrahousehold allocation. Section 3 presents the data. Sections 4 and 5 present the main empirical results and robustness checks. Section 6 analyzes the relative effectiveness of educational expenditure for children of different endowment, testing the hypothesis that parental allocation decisions are driven by the desire to maximize returns to educational spending. The final section concludes.

3.2 Modeling intrahousehold allocation of education

In order to provide a conceptual framework for the empirical analysis, this section outlines a simple model of parental decision-making. Assume that parents have two children and obtain utility from the welfare of each child as measured by his or her expected lifetime income. The parental utility function is a weighted sum of total income earned by both children and income earned by the poorer child, with the latter term capturing a parental preference for equality. The parameters θ_α and θ_β index the relative importance to parental utility of total income accruing to their family and the income accruing to the worst-off child, respectively.

$$U^i = \theta_\alpha(Y_1 + Y_2) + \theta_\beta \min \ \{Y_1, Y_2\} \tag{3.1}$$

Parents determine their children's expected lifetime earnings by allocating funds for education, provided at a price P_e presumed to be equal across children. Income is assumed to be an additively separable function of both education and endowment (denoted W) increasing in both arguments, such that $\frac{\partial Y}{\partial E \partial W} = 0$, $\frac{\partial Y}{\partial E} > 0$ and $\frac{\partial Y}{\partial W} > 0$.

$$Y_i = f(E_i) + g(W_i) \tag{3.2}$$

In this framework, the parents' optimization problem conditional on previous investments can be written as follows:

$$max \ U(Y_1(E_1, W_1), Y_2(E_2, W_2)) \ \ s.t. \tag{3.3}$$

$$\begin{aligned} (E_1 + E_2)P_e &\leq W & (3.4)\\ Y_i &= f(E_i, W_i) & (3.5)\\ E_i &\geq & (3.6) \end{aligned}$$

3.2.1 Parental allocations in an investment model

Assume that $\theta_\beta = 0$ and thus parents seek to maximize only total income earned by both children. The solution is described by the following first-order conditions.

$$\frac{\partial Y_1}{\partial E_1} - \lambda + \mu_1 = 0 \tag{3.7}$$

$$\frac{\partial Y_2}{\partial E_2} - \lambda + \mu_2 = 0 \tag{3.8}$$

$$W - E_1 - E_2 \geq 0, \lambda \geq 0 \tag{3.9}$$

$$E_1 \geq 0, \mu_1 \geq 0 \tag{3.10}$$

$$E_2 \geq 0, \mu_2 \geq 0 \tag{3.11}$$

Given the assumption that Y_i is strictly increasing in E_i, the budget constraint will always be binding; the non-negativity constraints may, however, be binding. If $\mu_i > 0$ and $\mu_j = 0$, then $E_i = 0$. From (3.7) and (3.8), $\frac{\partial Y_j}{\partial E_j} > \frac{\partial Y_i}{\partial E_i}$, with $E_j > E_i = 0$. Thus corner solutions obtain when returns to education are convex, and one randomly chosen child receives all the education.

If the condition for a corner solution is not satisfied, the interior solution is defined by the first-order condition $\frac{\partial Y_1}{\partial E_1} = \frac{\partial Y_2}{\partial E_2}$, and this holds only when returns to education are concave. Given the assumption that returns to education are equal across children, optimization given concave returns entails equal allocations of education. If returns to education are linear, any allocation of education equalizes returns and satisfies the condition for optimality.

3.2.2 Parental allocations in an equality-preferring model

Now assume that $\theta_\alpha = 0$ and parents seek only to maximize income earned by the poorer child. Utility is maximized where $Y_1 = Y_2$. If the children are equally able, $E_1 = E_2$. If child 1 is assumed to be more able, satisfaction of the condition of equality in incomes requires $E_2 > E_1$ in order to counterbalance the higher expected income of the more able child.

3.2.3 Parental allocations in a hybrid model

Now assume that $\theta_\alpha > 0$ and $\theta_\beta > 0$. In this case, the outcome is indeterminate. The interior solution, if it exists, is defined by the following first-order condition.

$$\theta_\alpha f'(E_1) = (\theta_\alpha + \theta_\beta) f'(E_2) \tag{3.12}$$

Here, E_1 is the allocation of education to the child with the greater endowment. There is also a corner solution in which all education is provided to a single child. The optimality of either solution is determined by the degree of convexity in returns to education and the relative magnitudes of the θ parameters indexing the preference for maximizing aggregate income versus equalizing income between the two children.

In the stylized modeling framework used here, a parental utility function that maximizes total income is consistent with full specialization in one child or equal allocations to both, while a pure parental preference for equality entails allocating more education to the child with the lower endowment. Intermediate preferences between the two extreme cases are compatible with a range of different allocations.

3.3 Data: Gansu Survey of Children and Families

The data set employed in this analysis is the Gansu Survey of Children and families (GSCF), a panel, multi-level study of rural children's welfare outcomes conducted in Gansu province, China. The first wave, conducted in 2000, surveyed a representative sample of 2000 children in 20 rural counties aged 9-12 as well as their mothers, household heads, teachers, principals, and village leaders. The second wave, implemented in 2004, supplemented the first wave with a sample of the younger siblings and fathers of the target children.

Gansu, located in northwest China, is one of the poorest and most rural provinces in China. In 2005, per capita income in rural areas was 1979.88 yuan, or less than $250; this was the second-lowest level of rural per capital income in China. 70% of the provincial population lived in rural areas (National Bureau of Statistics of the PRC 2006). Fertility rates are relatively low, with the average size of a family household 3.97 persons (UNESCAP 2002). The decline in fertility has generated an extremely high male-to-female sex ratio as traditional preferences for a son have led to increased sex-selective abortion, female abandonment and the underreporting of girls. The sex ratio in Gansu in 2000 was 111.2, close to the national average of 113.6 (Banister 2004).

This analysis will focus on a subsample of the families in the survey: those with two observed children. Here, the child aged 9-12 identified in the first round of the survey is referred to as the index child; in families where the index child had a younger sibling of school age, that child was surveyed in the second round. If these two children are the only children in the household, this constitutes a complete survey of parental allocations and child endowment, and these households are included in the analysis. Such complete data is available on 579 families, and they constitute the relevant subsample.[1]

Panel A of Table 3.1 compares the demographic characteristics of the subsample and the overall sample; no statistically significant difference is apparent in net income, per capita income, or parental education. The only significant difference between the two samples is in parental age: parents in the subsample are younger, reflecting the exclusion of households with larger numbers of children who will generally be headed by older parents.

The dependent variable of interest is educational expenditure per child per semester, reported by the mother in six categories: tuition, educational supplies, food consumed in school, transportation and housing, tutoring and other fees. In the Chinese educational context, supplies, tutoring, and other fees correspond to expenditure undertaken by the household to improve a child's academic performance, independent of the school attended. Expenses for transportation, housing and food, on the other hand, may also vary in accordance with the choice of school, and particularly the choice to have a child board at school or not. Discretionary expenditure is defined as the sum of all expenditure excluding tuition. Summary statistics for average expenditure per child for the subsample of families analyzed can be found in panel B of Table 3.1. Total educational expenditure averages slightly less than 300 yuan per child per semester, suggesting a total of 1160 yuan for two children over a year. This indicates that an average of 15% of mean household income is allocated to educational expenses.

The measurement of the child's endowment is height-for-age, normalized to a Z-score using the World Health Organization growth charts for children of ages 2-18. Height-for-age is widely used in the literature as a measure of endowment and a summary indicator of physical robustness, and it is correlated with a range of physical and cognitive indicators (World Health Organization 1995, Grantham-McGregor, Cheung, Cueto, Glewwe, Richter & Strupp 2007). At the same time, evidence suggests it largely reflects the history of nutrition or health prior to age three, as after this age catch-up for a child stunted in infancy is limited (Martorell 1999, Alderman, Hoddinott & Kinsey 2006). Accordingly, a robust relationship between height-for-age and early childhood shocks is expected. Summary statistics on height-for-age in the sample and the subsample are also shown in Table 3.1.

The primary data is supplemented by climatic data for Gansu. Grain yield data pre-

[1]The average number of children in households in this sample is 2.2. While the One-Child Policy was in effect during the period in which these children were born, many rural households could nonetheless have two children legally under various exemptions to the policy (Baochang, Feng, Zhigang & Erli 2007). Other households may simply have defied the rules. It is not possible using this dataset to accurately identify for each household whether it was in technical compliance with the policy.

1996 is from data tabulated by the Ministry of Agriculture; grain yield post-1996 is drawn from annual editions of the Rural Chinese Statistical Yearbook. Grain yield is measured annually at the county level in tons per hectare. Rainfall data is from the data collected by the Carbon Dioxide Information Analysis Center and is measured in mean millimeters per month. Data at the station level is interpolated to generate county-level measures using the inverse distance weighting method.[2]

3.4 Empirical evidence

3.4.1 Ordinary least squares

The primary relationship of interest in this analysis is equation (3.13), where the dependent variable is reported household expenditure on the education of child i in grade level g, household h and born in year t, denoted E_{ight}. The independent variable is endowment as measured by height-for-age, denoted H_{ight}. η_h is a household fixed effect that absorbs any household-level heterogeneity in the propensity for educational spending, and λ_g is a grade fixed effect that absorbs variation by grade in expenditure.

Because the subsample is composed of two-child families, a household fixed effects specification is equivalent to estimation of the equation in first differences across the two children. $\gamma_{t,elder}$ is a year-of-birth dummy for the elder child in each household, which serves to normalize the difference in height-for-age between the children relative to the mean difference among all households with a first child born in the same year. ϵ_{ight} is a child-specific error term. Standard errors are clustered by county and the year of birth of the second child, to allow for arbitrary correlation in the difference in height-for-age across different households in the same county and with a second child born in the same year. The equation of interest is thus the following:

$$E_{ight} = \beta H_{ight} + \eta_h + \lambda_g + \gamma_{t,elder} + \epsilon_{ight} \tag{3.13}$$

The equation is estimated for each of the six categories of educational expenditure, as well as for total discretionary expenditure and a dummy dependent variable for enrollment. The results, shown in Table 3.2, are insignificant. However, there is the potential for bias in these results if endowment measured at the age of primary school already embodies a

[2]Using this method, each county's measurement is calculated as the weighted average of measurements at stations within 100 kilometers of distance; the weights are equal to the distance between the station and the county divided by the sum of distances.

significant component of prior parental investment. The child who has already been the target of greater parental investment will appear to have a greater endowment, and, if there is some serial correlation in parental behavior, is likely to continue to receive more substantial investments. This will generate an upward bias in the estimated coefficients. Eliminating this bias is the goal of the identification strategy.

3.4.2 First stage across households

The key to identification in this case is the use of a climatic, and thus nutritional, shock that is correlated with the relative endowment of the two children. The climatic measures of interest are grain yield and rainfall. Accordingly, the first stage to be estimated is the following, where S_{ight} denotes the climatic shock for child i and γ_t is again a year-of-birth fixed effect.

$$H_{ight} = \beta S_{ight} + \gamma_t + \epsilon_{ight} \tag{3.14}$$

Due to soil erosion and a pattern of heavy rainfall during the harvest months (June and July) that is highly damaging, rainfall in this region is generally negatively correlated with grain yield (Cook, Fengrui & Huilan 2000), a relationship that is evident in the data employed here. Given this negative correlation, it is expected that rainfall will have an impact on early child nutrition and thus height-for-age that is the opposite of the effect for grain yield.

Table 3.3 presents the coefficients for the first stage, regressing height-for-age on climatic shocks in utero (defined as the calendar year prior to the year of birth) and in infancy (defined as the year of birth).[3] The first panel displays results employing rainfall as a measure of climatic shocks, and the second panel presents results employing grain yield. The first and second columns show the results of regressing height-for-age in the entire sample of children measured. Columns 3 through 6 show parallel regressions for a sample limited to the older child (child 1) and the younger child (child 2) in two-child households, to confirm that the relationship holds for both subsamples.

All specifications are estimated both with and without year-of-birth fixed effects and employing two-way clustering by county and year, allowing for both serial and spatial corre-

[3]Given that grain yield is reported only by year, it is not possible to adjust estimates of grain yield in infancy to account for variation in the month of birth. However, rainfall is reported monthly, and accordingly estimates of rainfall in the first twelve months of life as well as the nine months of gestation can be generated employing the reported month of birth. Using these alternate measures leads to no significant changes in the first stage coefficients.

lation. Though the instrument employed in the subsequent analysis will be climatic shocks in infancy as this relationship is more robust, the first stage for both set of shocks is shown to demonstrate the consistent pattern of such effects over the critical period of cognitive and physical development.

The results show coefficients on rainfall that are negative and significant, a result consistent with the negative impact of rainfall on grain yield already analyzed. The coefficients on grain yield are positive and significant, as expected: higher grain yield corresponds to greater nutritional availability in infancy, resulting in greater height-for-age and increased cognitive ability.

It is also informative to analyze whether the relationship between climatic shocks and endowment is similar for both the first-born and second-born children. As evident in Table 3.3, the first stage employing grain yield is positive and narrowly insignificant for the first child, and then much larger and highly significant for the second child. This reflects a sharp climatic shift that occurred in a subset of the sample counties lying in the Heihe river basin. Though these counties were consistently characterized by much higher mean grain yield over this period, they also experienced very sharp increases in yield post-1990 following the reversal of a previous process of desertification. This generates a much more robust first stage among children born post-1990. In the case of rainfall, the hypothesis of differing coefficients between the older and younger siblings can be rejected when a full set of cohort fixed effects are included. The hypothesis of different coefficients in the first stage for male and female children among either the older or younger siblings can be rejected for both instruments.

3.4.3 First stage with household fixed effects

Shifting the focus to a fixed-household, sibling-difference framework, Column (1) of Table 3.4 displays the results from the first stage estimated with household fixed effects and year-of-birth fixed effects for the older child, analogous to the ordinary least squares specification. Standard errors are clustered by county and the year of birth of the second child. The equation of interest is the following:

$$H_{ight} = \beta S_{ight} + \eta_h + \gamma_{t,elder} + \epsilon_{ight} \tag{3.15}$$

The results show a positive and significant correlation between the within-household difference in grain yield in the county and year of birth and the observed difference in height-for-age. Similarly, there is a negative and significant relationship between the difference in rainfall and the difference in height-for-age. When a child is born in a year with relatively favorable climatic shocks compared to his or her sibling, this child is observed to have relatively greater height-for-age.

As a robustness check, I also examine whether there is evidence of cross-dependence of shocks: controlling for his or her own shocks in infancy, the exclusion restriction requires that there is no dependence of one sibling's height-for-age on shocks experienced during the infancy of the other sibling. For example, one major threat to the identification strategy would be reallocation of resources by households in response to an adverse event: e.g., following the birth of the second child, households could preferentially direct resources to either the first or the second child when a negative shock occurs. This would be evident in a significant relationship between the older child's height and the shock to the younger child.

In order to test for this form of cross-dependence, I estimate the following equation, where S_{jght} denotes the climatic shock for sibling j and year-of-birth fixed effects are again included.

$$H_{ight} = \beta_1 S_{ight} + \beta_2 S_{jght} + \gamma_{t,elder} + \epsilon_{ight} \tag{3.16}$$

I estimate this equation separately for first-born children and second-born children; columns 2 and 3 of Table 3.4 report the results, suggesting that there is no relationship between endowment and the climatic shocks experienced during a sibling's infancy. This evidence indicates that the primary channel for the impact of early childhood shocks on height-for-age is via the effect on a child's own physical development, not a household-level mechanism that would induce cross-dependence of one sibling's outcome on another's shock. Given the absence of any such mechanism, the identification strategy remains plausible.

The magnitude of the coefficients indicates that a one standard deviation increase in grain yield in the county in the year of birth for the older sibling, holding the younger sibling's shock constant, will increase the difference in height-for-age between them by .153. This is equivalent to 13% of the mean height-for-age in levels, and about 200% of the mean difference in height-for-age. In other words, the mean difference in height between the older and younger siblings would be eliminated if there were a counterfactual increase in grain yield in the older child's year of birth corresponding to one half of the year-on-year standard deviation in that county. Using the first stage in rainfall, a one standard deviation increase in rainfall for the older child will increase the difference in height for age by .22, a marginally larger effect. The similarity in these estimates suggests they reflect the same fundamental relationship between climatic shocks and nutrition.

The F-statistics in the first stage suggest that rainfall has considerably greater predictive power than grain yield. This could be due to mismeasurement in official grain yield statistics, or an imperfect correlation between grain yield and the returns on other crops or other household income-generating activities. While there is little risk of bias due to weak instruments in the regressions employing rainfall, the results employing grain yield may suffer from limited precision due to the weaker first stage.

3.4.4 Reduced form and two-stage least squares

Table 3.5 shows the reduced form results of regressing the dependent variables on the climatic shocks. The equation estimated is the following, where again η_h, λ_g and $\gamma_{t,elder}$ denote household fixed effects, grade fixed effects and year-of-birth fixed effects for the older sibling.

$$E_{ight} = \beta S_{ight} + \eta_h + \lambda_g + \gamma_{t,elder} + \epsilon_{ight} \tag{3.17}$$

The first rows in Panel A and Panel B show the results for the preferred basic specification including year of birth fixed effects for the elder child; standard errors are again clustered by county and the year of birth of the second child. The second row shows the same effects with grade fixed effects included, and the third and fourth row of each panel converts four spending categories (transport/housing, food, tutoring and other) to dummy variables, given that a large number of zeros are observed in those categories. The results show a significant and positive relationship for rainfall shocks, and a significant and negative relationship for grain yield shocks: in both cases, educational spending favors the child who has been subject to a negative shock in infancy.

In the instrumental variables specification, equation (3.13) (reproduced here) is estimated employing the climatic shocks as an instrument for height-for-age H_{ight}. Results are again reported with and without grade fixed effects.

$$E_{ight} = \beta H_{ight} + \eta_h + \lambda_g + \gamma_{t,elder} + \epsilon_{ight}$$

The results reported in Table 3.6 are clear and consistent. In Panel A, the coefficient on height-for-age is negative and significant, indicating that children with a greater endowment receive less educational expenditure, for discretionary expenditure, transportation/housing and food in the preferred expenditure employing grade fixed effects. In Panel B, the point estimates are larger but imprecisely estimated. However, discretionary expenditure, food and tutoring are close to significant at conventional levels even employing grade fixed effects (p-values between .1 and .15).

The only significant impact on enrollment is found employing rainfall as an instrument in absence of grade fixed effects, suggesting that children with greater height-for-age are less likely to be enrolled in school. However, this effect disappears when grade fixed effects are added, suggesting that this reflects differences in probability of enrollment across grades (particularly, a lower probability of enrollment at older ages), rather than any systematic within-grade difference.[4]

The inclusion of grade fixed effects may be troubling given the potential objection that grade level in itself represents an outcome reflective of past parental investment: children who have already benefited from greater parental investments may have advanced to a higher grade level. The extremely low magnitude of the coefficient on enrollment, however, suggests that the enrollment decision is not a significant margin of parental decision-making that would induce large endowment-based variation in grade level. There is little evidence to indicate that the inclusion of grade fixed effects biases the conclusions, particularly in light of the consistency of the results.

To interpret the magnitude of the coefficients, consider that the average difference in height-for-age between siblings is -.07 of a standard deviation. The estimates using rainfall suggest that in response to such a difference, parents would redirect 7% of mean individual discretionary expenditure toward the weaker child; the estimates using grain yield indicate that parents would redirect 22% of expenditure.[5] Though relatively small in a given semester, such differences in expenditure are evident across sibling pairs in a variety of grade levels, suggesting that the cumulative impact if parents persist in such compensatory behavior could be large.

The difference in magnitude between the estimates using the two shocks is unsurprising given the evidence previously presented about geographical variation in the first stage relationship between climatic shocks and height-for-age. The complier households in an instrumental variable specification employing grain yield, those in which a difference in endowment between the two children is induced by shifts in relative grain yield in their birth years, are disproportionately drawn from the river-basin counties experiencing dramatic increases in yield. These counties were already characterized by higher agricultural yields and thus greater income. The estimated response in educational spending is thus much larger. The coefficient on grain yield corresponds to a 9% shift employing discretionary spending

[4]Overall rates of school enrollment in this population are extremely high. 98% of second-born children are reported enrolled in school and 91% of first-born children remain enrolled in school, with some dropout at older ages.

[5]These estimates correspond to the coefficients estimated in the absence of grade fixed effects.

in the river basin counties as the reference mean, an effect of magnitude comparable to the estimated effect using rainfall.

This evidence suggests the hypothesis that parental allocations of education are independent of children's endowment can be rejected. Parental allocations favor the child with lower endowment, though the hypothesis of full specialization in either child can also be rejected. In the modeling framework outlined, the observed pattern of parental allocations of educational expenditure is consistent with a parental preference for equality among children. This requires the use of education as a compensatory device for differences in endowment, as observed in the greater allocations to children with lower observed height-for-age.

In this dataset, there are only two other types of parental investments that are reported disaggregated by child. The first is medical attention (more specifically, the number of visits to a doctor), analyzed in the robustness checks below; the second is time spent by the mother assisting each child with his or her homework. In 80% of households, the mother reports spending no time assisting with homework for either child. Accordingly, the dependent variable of interest is defined as a dummy equal to one if the mother spends any time assisting a child with homework. There is weak evidence in the two-stage least squares specification with household fixed effects of the same pattern in which investment of maternal time favors the weaker child, but the results are not statistically significant at standard levels.

3.5 Robustness checks

3.5.1 Sibling parity effects

Given the difference in mean height-for-age between the older and the younger children, one potential concern is that height-for-age may simply proxy for birth order, and that a pattern of expenditure favoring the weaker child may reflect a pattern of expenditure favoring the first-born child. It is not possible to test for sibling parity effects directly using the household fixed effects specification. In order to address this point, the main equation (3.13) could be re-estimated in a cross-section, thereby eliminating the household fixed effects, and including both own height and sibling's height as independent variables. H_{jght} denotes the sibling's height; both measures of height are instrumented with the respective shocks in infancy. ρ_c denotes a county fixed effect, and ζ_h denotes the household-specific error component.

$$E_{ight} = \beta_1 H_{ight} + \beta_2 H_{jght} + \beta_3 P_{ight} + \gamma_t + \rho_c + \zeta_h + \epsilon_{ight} \qquad (3.18)$$

In order to estimate unbiased coefficients on H_i and H_j, this equation requires a stronger exclusion restriction than the household fixed effects specification: the climatic shock employed to instrument for each sibling's height must be orthogonal not only to the unobserved residual for that child, but also to the residual for the household. In particular, climatic shocks must have no impact on the household's overall budget constraint conditional on county and year fixed effects. However, the cross-sectional specification has the advantage of allowing the direct implementation of a test for the presence of sibling parity effects by including a dummy variable for sibling parity, denoted P_{ight}. If height-for-age is merely an imperfect proxy for sibling parity, only parity will be a significant predictor of expenditure, and height-for-age should be rendered insignificant.

The results of estimating (3.18) are shown in Table 3.7, employing rainfall as an instrument in Panel A and grain yield in Panel B; all specifications include both year-of-birth and grade level dummies. Standard errors in this case are clustered by county and the year of birth of the sibling. The results are largely consistent with the previous specification despite the known source of bias. Own height is generally insignificant, and sibling height is positive and significant for five categories of expenditure, indicating that educational expenditure increases when a child has lower height-for-age than his or her sibling. More importantly, the uniform insignificance of the parity dummy suggests that there is no significant variation in expenditure as a product of birth order. The only exception is a negative coefficient on sibling parity in Column (3) of Panel B.

Additional tests can be implemented by employing the data on medical expenditure previously described. Data on medical expenditure on children is available from two sources. First, the mother reports the number of visits to a doctor or to a clinic or hospital for each child, as well as the number of medical episodes and days of school missed due to sickness for each child. Second, the head of household (normally the father) separately reports medical expenses for each child and the number of days ill over the last month.

In order to test the hypothesis of sibling parity, parental reports of expenditure on a child's health (doctor visits, clinic visits and medical expenditure), denoted E_{iht}, are regressed on the indicator of medical necessity reported by that parent M_{iht}. For the mother's reports, this is medical episodes or days of school missed due to illness; for the father's reports, this is days ill over the past month. Sibling parity P_{iht} is also included, along with year-of-birth fixed effects γ_t.

$$E_{iht} = \beta_1 M_{iht} + \beta_2 P_{iht} + \gamma_t + \epsilon_{iht} \tag{3.19}$$

Table 3.8 shows these results, and, again, sibling parity is found to be uniformly insignificant, indicating no pattern of systematic favoritism by birth order.[6] The fact that this is corroborated by reports from both parents is encouraging, as one might hypothesize that if one parent systematically favors the older or younger child, he or she might over-report his or her episodes of illness in order to justify additional expenditure on that child. This would generate upward bias on the estimated coefficient on M_{iht}, and potentially obscure a pattern of favoritism by sibling parity. However, the coincidence of the estimated results from both parents suggests that reporting bias is unlikely to be a factor biasing the results.

3.5.2 Gender effects

A second potential confounding factor is gender. Given the evidence from other sources of gender bias in household decision-making in China, the effect of gender on parental allocations may outweigh any observed effect for endowment. On the other hand, simply controlling for gender is unlikely to be satisfactory. Given the abundant anthropological and demographic evidence on abortion, abandonment, or underreporting of female children in China (Coale & Banister 1994, Qian 1997), it is implausible to assume that the gender of both children can be assumed to be random. Households with different gender balances among their children are likely to differ materially along other observable and unobservable dimensions.

However, in this case, the gender of the first child is a plausibly exogenous observation, as anthropological evidence indicates that selection for gender occurs principally in births subsequent to a first-born daughter and that selective abortion prior to the birth of a first child is unusual (Baochang et al. 2007, Banister 2004). The evidence in this sample is consistent with this hypothesis. The sex ratio for the first child in this sample is not significantly different from .5, while for the second-born child, the sex ratio is highly imbalanced.

The observations on the gender of the first and second child are thus assumed to constitute one exogenous observation (the gender of the first child) and one endogenous observation (the gender of the second child). In a family fixed-effects framework such as that employed here, the relevant dependent variable is the difference in gender between the children, which is likewise endogenous. However, the gender of the first child is an appropriate instrument if the assumption of exogeneity is plausible. The main equation of interest is again estimated

[6] There is little evidence that medical expenditure follows a compensatory pattern based on height-for-age after controlling for the number of acute medical episodes experienced by each child. While the estimated coefficient is negative and thus consistent with prior analysis, it is not statistically significant.

employing climatic shocks as instruments, and using the gender of the first child as an instrument for the difference in gender, yielding the following estimating equation.

$$E_{ight} = \beta_1 H_{ight} + \beta_2 G_{ight} + \eta_h + \lambda_g + \gamma_{t,elder} + \epsilon_{ight} \tag{3.20}$$

The estimation results are shown in Table 3.9. The coefficient on endowment is again significant and negative for discretionary expenditure. Gender, perhaps surprisingly, proves to be uniformly insignificant with the exception of a positive coefficient on enrollment in the specifications lacking grade fixed effects and on tuition when employing grain yield. However, this effect disappears when grade fixed effects are added. Conditional on endowment, discretionary allocations of educational funding between boys and girls appear to be similar. There is only minimal evidence of preferential allocations to male children.

3.5.3 Selection bias

Selection into the sample of two-child families observed in this analysis would also constitute a violation of the exclusion restriction. If families with certain characteristics are more or less liekly to suffer an adverse mortality event as the result of the same climatic shock, then the pattern of shocks may affect the ultimate pattern of allocations by determining the surviving number of children, and hence inclusion or exclusion in the sample.

Due to the absence of complete data on retrospective familial mortality, it is not possible to directly examine child mortality as a function of varying climatic shocks. An alternative strategy to test for selection effects exploits the presence of extremely severe climatic shocks that are most likely to be associated with increased mortality. If there is selective survival among children born in those years, this would be expected to produce an attenuation toward zero in an otherwise negative relationship between climatic shocks and child mortality. This reflects the fact that surviving children, while weakened by adverse conditions in infancy, are nonetheless likely to be genetically more robust and thus have a propensity toward greater health. On the other hand, if selection via differential mortality is not an important phenomenon, there should not be an attenuation of the relationship between shocks and health outcomes as the severity of the shock increases.

To test this hypothesis, height-for-age in the pooled sample of older and younger siblings is regressed on the climatic shock variable and four dummy variables indexing negative shocks of different severity in the first year of life. A severe shock is defined as a year in which rainfall (grain yield) in that county was above (below) the 10th percentile over all counties in the years of interest; a negative shock is a year in which rainfall (grain yield) was above (below) the 25th percentile. The regressions are estimated with cohort fixed effects and county-year clustering. If selective mortality is a significant phenomenon, the coefficients on these dummies are expected to be significant and positive: the relationship between rainfall and height-for-age is less negative for households experiencing a particularly severe shock.

The results, shown in Table 3.10, show coefficients on the dummy variables that are uniformly insignificant. This indicates that there is no attenuation in the relationship between climate shocks and height-for-age for extremely severe shocks. Selective mortality thus appears to be minimal, and selection into the observed subsample of two-children family is unlikely to be a major source of bias.

3.6 Returns to educational expenditure and endowment

These results raise the question of whether the observed pattern favoring the weaker child exhibited is a compensatory response intended to provide consumption-like educational benefits to children with lower endowments, or whether this allocation strategy reflects differential returns to educational expenditure for children of differing levels of endowment. If, for example, educational expenditure has higher returns for the child with a lower endowment, then the observed strategy could be interpreted as maximizing returns to educational investment.

In order to test this hypothesis, the panel data of observations on the older sibling is employed to evaluate gains in academic attainment over time as a function of height-for-age, again instrumented with climatic shocks, educational expenditure and the interaction between the two. The equation of interest is (3.21), where ΔT_{ight} denotes the gain in academic skills as measured by the difference in scores in the first and second survey waves on grade-specific academic tests in Chinese and mathematics. E_{ight} denotes educational expenditure, also interacted with height-for-age. T^1_{ight}, the test score on the first round, is included as an independent variable, though not reported, in order to account for mean reversion. The test scores are normalized by grade to have mean zero and standard deviation one.

$$\Delta T_{ight} = \beta_1 H_{ight} + \beta_2 E_{ight} + \beta_3 H_{ight} \times E_{ight} + \beta_4 T^1_{ight} + \gamma_t + \zeta_h + \epsilon_{ight} \tag{3.21}$$

Because the younger sibling was not included in the first wave of testing due to their extremely young age, this equation can only be estimated in the cross-section without household fixed effects. This, again, raises the concern of income effects that violate the exclusion restriction. The year-of-birth dummies will partially absorb the effect of climatic shocks on income. However, the residual variation in income will lead to upward bias in β_1, as those children with higher height-for-age are likely to be from households with higher income and thus higher educational expenditure.

The endogeneity of expenditure is another source of bias in this equation: households that spend more on education are likely to encourage educational achievement in other ways, creating upward bias in β_2. Given the evidence presented on the importance of compensating behavior, it is a reasonable prior to expect downward bias on β_3 as well. Households that have higher educational spending and value education are expected to also be more effective in encouraging catch-up by children with a weaker endowment, and this will generate spurious evidence of higher returns to educational spending on those children.

Despite this source of bias, however, the results in Table 3.11 show that the interaction effect is generally insignificant, though it is positive and highly significant for tutoring in specifications employing both grain yield and rainfall, and positive and marginally significant for total expenditure and discretionary expenditure for specifications employing grain yield. This indicates that while there is little variation in the relative effectiveness of educational expenditure for children of different endowments, the interaction effect is positive when it is significant. Returns to at least one type of educational expenditure are increasing in endowment, and these results may be biased downwards (i.e., the true effect may be even larger). For this reason, the hypothesis that parents preferentially direct expenditure to children of lower endowments in order to maximize the total returns on their educational spending seems implausible.

Additional evidence on this point can be provided by restricting the sample to families whose two children include at least one girl. In this case, it is arguably less plausible that households are seeking simply to maximize the return on their educational investment in terms of scholastic achievement and ultimate income-generating potential, given the lower rates of female labor force participation. While the estimated coefficients in this specification are less precise, perhaps unsurprising given the smaller size of the relevant subsample, they remain close to significant at conventional levels (p-values between .1 and .2) and of comparable magnitude; the hypothesis that the effects are identical cannot be rejected.[7] This is consistent with the previous evidence that the observed pattern of educational expenditure

[7]Tabulations available on request.

does not reflect a strategy to maximize returns to expenditure.

3.7 Conclusion

In the previous literature on intrahousehold allocation, the question of the presence or absence of family aversion to inequality has received extensive analytical attention. However, little evidence has been presented regarding the nature of parental responses to systematic differences in endowment among children, particularly among children who are not twins. This paper analyzes the relationship between parental allocations of educational expenditure and endowment, testing a model in which parents seek to maximize the total returns to all educational investments against one in which they have a preference for equality of expected lifetime income across their offspring.

Employing an identification strategy that relies on the correlation between climatic variation and children's endowment, mediated through the impact of nutritional shocks in infancy on physical development, I find a pattern of preferential allocations of discretionary educational expenditure to child of lower endowment, conditional on the attained grade level. This is consistent with a parental preference for equality of outcomes. The relationship is robust across multiple specifications and measures of expenditure, and robust to the inclusion of both gender and sibling parity.

However, evidence suggests that this pattern of allocations is not a response to higher returns to educational expenditure for children of lower endowment. In fact, for some categories of expenditure, returns are higher for children with greater height-for-age, indicating that households seeking to maximize returns to their educational investment would invest in the more robust child.

These results suggest that, at least in the area of education, the household is serving as a mechanism for the mitigation of existing inequalities. It is impossible to rule out the possibility that the observed allocative decisions would be altered by an external shock. However, based on this evidence it seems a priori plausible that a positive income shock to the households examined would disproportionately benefit the more vulnerable among the children, where vulnerability is measured as a lower physical endowment. For the purposes of the welfare analysis of potential household interventions, this is an encouraging result that suggests that household-level interventions will in fact improve welfare outcomes for the weakest members of a family.

3.8 Tables

Table 3.1: Summary statistics

Panel A: Demographic data

	Sample	Subsample
Net income	7297.35	7459.49
Income per capita	1825.92	1850.27
Father educ.	6.84	6.84
Mother educ.	4.17	4.17
Father age	38.6	35.1
Mother age	35.2	33.0
Index child age	11.0	11.1
Height-for-age Z-score	-1.18	-1.20
Obs.	1918	596

Panel B: Educational expenditure per child

	Mean	Std. Dev.	Max.
Total	291.78	384.00	7760
Discretionary	112.32	205.50	2240
Tuition	179.47	228.56	6000
Supplies	39.43	43.39	600
Transport / Housing	13.16	48.86	600
Food	42.3	127.99	1700
Tutoring	7.20	35.17	1000
Other fees	10.21	30.78	420
Obs.	596	596	596

Notes: Income is reported in yuan; educational expenditure is reported in yuan per semester.

Table 3.2: OLS

	Enrollment (1)	Discretionary (2)	Tuition (3)	Supplies (4)	Transport (5)	Food (6)	Tutoring (7)	Other (8)
Height-for-age	.013	-3.933	6.990	-.516	.151	-4.177	1.136	.126
	(.010)	(9.083)	(5.743)	(.611)	(1.602)	(7.084)	(.704)	(.252)
+ Grade FE	7.55e-17	-6.873	-.168	-.991	-2.232	-4.696	.882	.165
	(6.83e-17)	(4.596)	(2.378)	(.822)	(1.380)	(2.938)	(.552)	(.322)
Mean (levels)	.93	94.31	170.27	36.50	6.73	9.24	6.73	9.24
Median (levels)	1	40	130	30	0	0	0	0
Mean (differences)	.07	89.07	65.17	17.11	17.60	50.81	6.8	7.3
Obs.	422	422	377	377	377	377	377	377

Notes: The dependent variable is educational expenditure per child in the specified category; the independent variable is height-for-age. All regressions include household fixed effects, year of birth fixed effects for the older child, and standard errors employing two-way clustering by county and the year of birth of the second child; grade fixed effects are dummy variables for the grade level attended in school. The mean in differences reports the mean intra-household difference in expenditure (between the first and second child) in the specified category. Asterisks denote significance at the 1, 5, and 10 percent levels.

Table 3.3: First stage across households

	Full sample		First-born children		Second-born children	
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Rainfall			
Rainfall in utero	-.021	-.021	-.020	-.020	-.029	-.026
	(.005)***	(.005)***	(.002)***	(.004)***	(.005)***	(.005)***
Rainfall	-.021	-.020	-.028	-.028	-.034	-.027
	(.005)***	(.006)***	(.008)***	(.008)***	(.006)***	(.005)***
			Panel B: Grain yield			
Grain yield in utero	.125	.106	.094	.093	.160	.092
	(.023)***	(.028)***	(.043)**	(.044)**	(.027)***	(.040)**
Grain yield year 1	.114	.098	.068	.072	.147	.085
	(.023)***	(.027)***	(.047)	(.048)	(.029)***	(.042)**
Year-of-birth fixed effects	No	Yes	No	Yes	No	Yes
Mean height-for-age	-1.19	-1.19	-1.27	-1.27	-1.17	-1.17
Num of obs.	2524	2524	705	705	759	759

Notes: The dependent variable is height-for-age Z-score; the independent variable is the specified climatic shock measured either in utero (the year before birth) or in the first year of life. In Columns (3) and (4), the sample is restricted to first-born children; in Columns (5) and (6), it is restricted to second-born children. All regressions include standard errors employing two-way clustering by county and year of birth; asterisks indicate significance at 1, 5 and 10 percent levels.

Table 3.4: First stage within households

	Height-for-age (1)	Height (first-born child) (2)	Height (second-born child) (3)
		Panel A: Rainfall	
Rainfall	-.043 (.006)***		
Rainfall (child 1)		-.038 (.009)***	.017 (.012)
Rainfall (child 2)		-.00005 (.006)	-.044 (.012)***
Household fixed effects	Yes	No	No
Obs.	418	418	419
F stat.	40.924	12.870	7.679
		Panel B: Grain yield	
Grain yield	.520 (.221)**		
Grain yield (child 1)		.233 (.140)*	-.443 (.346)
Grain yield (child 2)		-.109 (.116)	.523 (.296)*
Household fixed effects	Yes	No	No
Obs.	418	418	419
F stat.	4.90	3.107	9.001

Notes: The dependent variable is height-for-age and the independent variables are climatic shocks (either rainfall or grain yield). In the first column, household fixed effects are included, and the equation is equivalent to a regression of the within-household difference in height-for-age on the within-household difference in climatic shocks. In the second and third columns, household fixed effects are excluded. In Column (2), height-for-age for the older child is regressed on climatic shocks observed in the infancy of both the first and second child; in Column (3), height-for-age for the younger child is regressed on both climatic shocks. All regressions include year-of-birth fixed effects for the older child and standard errors employing two-way clustering by county and year of birth of the younger child; asterisks indicate significance at 1, 5 and 10 percent levels.

Table 3.5: Reduced form

	Enroll. (1)	Discret. (2)	Tuition (3)	Supplies (4)	Trans./Housing (5)	Food (6)	Tutoring (7)	Other (8)
				Panel A: Rainfall				
Rainfall	.003	4.098	1.813	.242	.794	3.497	-.061	.112
	(.002)**	(2.517)	(.544)***	(.234)	(.496)	(1.961)*	(.190)	(.193)
+ Grade FE	-7.93e-20	2.798	.954	.115	.448	2.384	-.139	-.010
	(1.29e-16)	(1.702)	(.698)	(.085)	(.306)	(1.308)*	(.140)	(.137)
Expend. dummies					.006	.006	.008	-.006
					(.004)*	(.004)*	(.003)**	(.004)
+ Grade FE					.007	.007	.010	-.004
					(.004)**	(.004)**	(.003)***	(.002)*
				Panel B: Grain yield				
Grain yield	.026	-168.500	-24.653	-12.276	-26.997	-109.667	-8.158	-11.401
	(.025)	(41.108)***	(12.322)**	(.000)	(7.470)***	(29.436)***	(2.998)***	(4.290)***
+ Grade FE	1.10e-17	-117.281	16.185	-5.279	-17.848	-77.287	-6.803	-10.064
	(1.18e-15)	(31.977)***	(9.497)*	(2.324)**	(5.898)***	(23.193)***	(2.603)***	(3.400)***
Expend. dummies					-.187	-.187	-.161	-.100
					(.074)**	(.074)**	(.061)***	(.053)*
+ Grade FE					-.153	-.153	-.082	-.089
					(.052)***	(.052)***	(.039)**	(.045)**
Mean (levels)	.93	94.31	170.27	36.50	6.73	9.24	6.73	9.24
Median (levels)	1	40	130	30	0	0	0	0
Mean (differences)	.07	89.07	65.17	17.11	17.60	50.81	6.8	7.3
Num of obs.	419	419	375	375	375	375	375	375

Notes: The dependent variable is educational expenditure per child in the specified category; the independent variable is the specified climatic shock. All regressions include household fixed effects, year of birth fixed effects for the older child, and standard errors employing two-way clustering by county and the year of birth of the second child; grade fixed effects are dummy variables for the grade level attended in school. The mean in differences reports the mean intra-household difference in expenditure (between the first child and the second) in the specified category. Asterisks denote significance at the 1, 5, and 10 percent levels.

Table 3.6: Two-stage least squares

	Enroll. (1)	Discretionary (2)	Tuition (3)	Supplies (4)	Trans./Housing (5)	Food (6)	Tutoring (7)	Other (8)
Panel A: Rainfall								
Height-for-age	-.056	-102.169	-41.496	-5.385	-17.694	-77.939	1.356	-2.506
	(.029)*	(36.198)***	(14.352)***	(4.036)	(8.001)**	(27.268)***	(4.483)	(3.111)
+ Grade FE	1.89e-18	-66.754	-22.756	-2.748	-10.688	-56.875	.233	3.323
	(3.08e-15)	(29.269)**	(21.527)	(1.843)	(6.799)	(23.318)**	(3.355)	(3.812)
Spending dummies					-.172	-.172	-.224	.105
					(.077)**	(.077)**	(.073)***	(.089)
+ Grade FE					-.178	-.178	-.227	.090
					(.062)***	(.062)***	(.000)	(.071)
Panel B: Grain yield								
Height-for-age	.032	-358.421	-52.440	-26.112	-57.427	-233.276	-17.353	-24.252
	(.047)	(171.801)**	(35.774)	(13.755)*	(36.677)	(99.778)**	(9.276)*	(14.953)
Height-for-age	2.70e-17	-288.743	39.848	-12.996	-43.941	-190.280	-24.777	-16.748
	(2.90e-15)	(194.191)	(25.606)	(10.249)	(36.436)	(125.393)	(15.758)	(11.515)
Spending dummies					-.384	-.384	-.334	-.215
					(.250)	(.250)	(.196)*	(.176)
+ Grade FE					-.377	-.377	-.203	-.218
					(.308)	(.308)	(.163)	(.190)
Mean (levels)	.93	94.31	170.27	36.50	6.73	9.24	6.73	9.24
Median (levels)	1	40	130	30	0	0	0	0
Mean (differences)	.07	89.07	65.17	17.11	17.60	50.81	6.8	7.3
Num of obs.	419	419	375	375	375	375	375	375

Notes: The dependent variable is educational expenditure per child in the specified category; the independent variable is height-for-age, instrumented by the specified climatic shock. All regressions include household fixed effects, year of birth fixed effects for the older child, and standard errors employing two-way clustering by county and the year of birth of the second child; grade fixed effects are dummy variables for the grade level attended in school. The mean in differences reports the mean intra-household difference in expenditure (between the first child and the second) in the specified category. Asterisks denote significance at the 1, 5, and 10 percent levels.

Table 3.7: Educational expenditure and sibling parity

	Enroll. (1)	Discretionary (2)	Tuition (3)	Supplies (4)	Trans./Housing (5)	Food (6)	Tutoring (7)	Other (8)
Panel A: Rainfall								
Height-for-age	.002	-8.027	19.966	8.313	-6.470	-23.198	5.359	7.973
	(.003)	(21.786)	(27.871)	(6.853)	(8.215)	(20.923)	(3.012)*	(5.912)
Sibling height	-.008	84.497	69.537	6.971	11.666	57.095	1.251	7.500
	(.009)	(29.376)***	(21.689)***	(4.334)	(6.551)*	(22.341)**	(3.551)	(3.028)**
Parity	.002	1.916	-13.439	-.391	.873	3.742	.121	-2.425
	(.002)	(19.470)	(15.474)	(2.890)	(3.454)	(12.130)	(.471)	(4.704)
Panel B: Grain yield								
Height-for-age	-.194	-75.999	-132.328	9.940	-10.368	-27.855	1.215	5.670
	(.096)**	(63.696)	(51.961)**	(11.142)	(27.757)	(48.570)	(8.943)	(11.149)
Sibling height-for-age	-.110	33.972	-37.682	8.680	5.782	49.802	-2.817	5.488
	(.100)	(64.246)	(75.537)	(5.706)	(20.446)	(55.112)	(5.661)	(9.873)
Sibling height-for-age	-.024	-20.921	-45.370	-3.400	-.049	-1.667	-1.689	-4.858
	(.033)	(18.046)	(13.494)***	(.000)	(7.382)	(17.804)	(2.402)	(3.794)
Mean	.93	94.31	170.27	36.50	6.73	9.24	6.73	9.24
Median	1	40	130	30	0	0	0	0
Obs.	796	796	790	754	754	754	754	754

Notes: The dependent variable is educational expenditure per child in the specified category. The independent variables are height-for-age and sibling height-for-age, instrumented by the specified climatic shock in the year of birth, and a variable capturing sibling parity. All regressions include year of birth fixed effects and standard errors employing two-way clustering by county and the sibling's year of birth. Asterisks denote significance at the 1, 5, and 10 percent levels.

Table 3.8: Medical expenditure and sibling parity

	Doctor visit (1)	Doctor visit (2)	Clinic visit (3)	Clinic visit (4)	Medical expenditure (5)
Med episodes	.119		.034		
	(.015)***		(.025)		
Days missed		-.026		.122	
		(.004)***		(.021)***	
Days sick					18.626
					(1.196)***
Parity	.013	-.012	.001	.008	1.977
	(.048)	(.049)	(.234)	(.230)	(15.921)
Mean	1.597	1.597	1.849	1.849	60.773
Median	2	2	0	0	10
Num of obs.	817	817	846	846	839

Notes: The dependent variables are various measures of medical attention received by the child; the independent variables are measures of ill-health and a dummy for sibling parity. All regressions include year of birth fixed effects. Due to the presence of extreme outliers in expenditure, the top 1% of observations are trimmed; asterisks indicate significance at 1, 5 and 10 percent level.

Table 3.9: Educational expenditure and gender

	Enroll. (1)	Discretionary (2)	Tuition (3)	Supplies (4)	Trans./Housing (5)	Food (6)	Tutoring (7)	Other (8)
			Panel A: Rainfall					
Height for age	.054	-98.971	-42.032	-5.329	-17.866	-78.333	1.351	-2.544
	(.029)*	(33.089)***	(13.331)***	(4.115)	(8.053)**	(27.539)***	(4.497)	(3.275)
Gender	.046	18.666	15.695	-2.124	6.566	15.065	.153	1.436
	(.006)***	(15.143)	(9.843)	(3.374)	(4.221)	(15.061)	(1.095)	(1.987)
+ Grade FE	-9.58e-18	-66.757	-22.747	-2.751	-10.686	-56.876	3.323	.233
	(3.18e-15)	(29.194)**	(20.061)	(2.028)	(6.761)	(23.309)**	(3.789)	(3.372)
Gender	-1.75e-18	-2.838	10.484	-3.776	1.613	-.493	-.363	.181
	(7.26e-16)	(12.668)	(10.998)	(2.567)	(1.835)	(11.816)	(1.127)	(2.488)
			Panel B: Grain yield					
Height for age	.023	-309.448	-57.865	-26.362	-59.872	-240.757	-17.992	-25.195
	(.057)	(106.836)***	(34.323)*	(13.991)*	(38.683)	(105.332)**	(9.652)*	(15.317)*
Gender	.050	58.687	30.146	1.389	13.589	41.570	3.554	5.239
	(.019)***	(43.186)	(13.189)**	(6.054)	(8.939)	(38.285)	(3.026)	(3.977)
+ Grade FE	1.56e-17	-296.726	39.236	-12.121	-45.831	-195.367	-17.527	-25.880
	(2.79e-15)	(200.339)	(27.055)	(10.214)	(38.309)	(129.514)	(11.707)	(16.225)
Gender	3.73e-18	24.532	1.881	-2.688	5.807	15.632	2.394	3.387
	(.000)	(37.145)	(15.326)	(3.205)	(5.400)	(26.252)	(1.605)	(2.885)
Mean (levels)	.93	94.31	170.27	36.50	6.73	9.24	6.73	9.24
Median (levels)	1	40	130	30	0	0	0	0
Mean (differences)	.07	89.07	65.17	17.11	17.60	50.81	6.8	7.3
Obs.	418	418	375	375	375	375	375	375

Notes: The dependent variable is educational expenditure per child in the specified category; the independent variable is height-for-age, instrumented by the specified climatic shock, and gender, instrumented by the gender of the elder child. All regressions include household fixed effects, year of birth fixed effects for the older child, and standard errors employing two-way clustering by county and the year of birth of the second child; grade fixed effects are dummy variables for the grade level attended in school. The mean in differences reports the mean intra-household difference in expenditure (between the first child and the second) in the specified category. Asterisks denote significance at the 1, 5, and 10 percent levels.

Table 3.10: Height-for-age and severe shocks

	Height-for-age (1)	Height-for-age (2)
	Panel A: Rainfall	
Rainfall	-.018 (.007)***	-.022 (.005)***
Negative shock	-.042 (.107)	
Severe shock		.080 (.070)
	Panel B: Grain yield	
Grain yield	.114 (.033)***	.099 (.029)***
Negative shock	.077 (.119)	
Severe shock		-.0009 (.026)
Obs.	2524	2524

Notes: The dependent variable is height-for-age, and the independent variables are the specified climatic shock (entering linearly) and dummy variables for negative and severe shocks, defined as rainfall (grain yield) that is above (below) the 75th (25th) percentile for a negative shock, or above (below) the 90th (10th) percentile for a severe shock. All regressions include year of birth fixed effects and two-way clustering by county and year of birth; asterisks indicate significance at 1, 5 and 10 percent levels.

Table 3.11: Returns to educational expenditure

	Total (1)	Discretionary (2)	Tuition/Supplies (3)	Food/Trans./Housing (4)	Tutoring (5)	Other (6)
			Panel A: Rainfall			
Height for age	-1.514	-.349	.095	-.425	-.281	-.351
	(28.818)	(.071)***	(3.239)	(.244)*	(.042)***	(.018)***
Expenditure	.009	-.00005	-.001	-.007	.053	.004
	(.196)	(.012)	(.014)	(.068)	(.013)***	(.002)*
Expenditure int.	.012	-.001	-.004	-.007	.038	.003
	(.284)	(.012)	(.028)	(.063)	(.009)***	(.002)
			Panel B: Grain yield			
Height for age	-.417	-.369	-.457	-.394	-.163	-.361
	(.107)***	(.025)***	(.133)***	(.039)***	(.095)*	(.046)***
Expenditure	.001	.003	.001	.002	.113	.003
	(.0004)***	(.003)	(.0005)**	(.004)	(.044)**	(.009)
Expenditure int.	.0005	.002	.0006	.002	.083	.001
	(.0007)	(.003)	(.001)	(.003)	(.034)**	(.008)
Mean	136.970	46.501	91.777	12.525	.360	14.282
Median	113	25	85	0	0	0
Obs.	1585	1585	1564	1564	1564	1564

Notes: The dependent variable is the gain in test scores on a grade-specific exam in Chinese and mathematics between a first and second round of testing; the first test was administered in 2000, and the second test in 2004. Test scores are normalized by grade level to have mean zero and standard deviation one. The independent variables are height-for-age, instrumented by the specified climate shock, expenditure (in the specified category) reported in the year in which the first test was administered, and the interaction of expenditure and height-for-age. The test score on the first test is also included as a regressor, though not reported. All regressions include cohort fixed effects and two-way clustering by county and year; asterisks indicate significance at 1, 5 and 10 percent levels.

Bibliography

Acemoglu, Daron, Simon Johnson & James Robinson. 2001. "Colonial origins of comparative development." *American Economic Review* 91(3):1369–1401.

Alderman, Harold, John Hoddinott & Bill Kinsey. 2006. "Long-term consequences of early childhood malnutrition." *Oxford Economic Papers* 58(3):450–474.

Anderson, Siwan, Patrick Francois & Ashok Kotwal. 2011. "One Kind of Democracy.".

Appu, P.S. 1996. *Land reforms in India: A survey of policy, legislation and implementation.* New Delhi: Vikas Publishing House.

Aziz, Abdul & Sudhir Krishna. 1983. *The political economy of land reforms.* New Delhi: Ashish.

Bandyopadhyay, D. 1986. "Land Reforms in India: An Analysis." *Economic and Political Weekly* .

Banerjee, Abhijit. 2003. Contracting constraints, credit markets and economic development. In *Advances in Economics and Econometrics: Theory and Applications*, ed. Mathias Dewatripont, Lars Peter Hansen & Stephen Turnovsky. Cambridge: Cambridge University Press pp. 1–46.

Banerjee, Abhijit & Lakshmi Iyer. 2005. "History, institutions, and economic performance: The legacy of colonial land tenure systems in India." *American Economic Review* 95(4):1190–1213.

Banerjee, Abhijit, Paul Gertler & Maitreesh Ghatak. 2002. "Empowerment and efficiency: Tenancy reform in West Bengal." *Journal of Political Economy* 110(2):239–280.

Banister, Judith. 2004. "Shortage of girls in China today." *Journal of Population Research* 21(1):19–45.

Baochang, Gu, Wang Feng, Guo Zhigang & Zhang Erli. 2007. "China's local and national fertility policies at the end of the twentieth century." *Population and Development Review* 33:129–147.

Bardhan, Pranab. 2011. "Industrialization and the land acquisition conundrum." *Development Outreach* 20.

Bardhan, Pranab, Michael Luca, Dilip Mookherjee & Francisco Pino. 2011. "Evolution of land distribution in West Bengal 1967-2004: Role of land reform and demographic changes.".

Behrman, Jere. 1988. "Intrahousehold allocation of nutrients in rural India: Are boys favored? Do parents exhibit inequality aversion?" *Oxford Economic Papers* 40:32–54.

Behrman, Jere & Anil Deolalikar. 1990. "Intrahousehold demand for nutrients in rural South India: individual estimates, fixed effects and permanent income." *Journal of Human Resources* 25(4):665–696.

Behrman, Jere, Robert Pollak & Paul Taubman. 1982. "Parental preferences and provision for progeny." *Journal of Political Economy* 90(1):52–73.

Behuria, N.C. 1997. *Land reforms legislation in Karnataka.* New Delhi: Sage.

Bergmann, Theodor. 1984. *Agrarian reform in India, with special reference to Kerala, Karnataka, Andhra Pradesh and West Bengal.* New Delhi: Agricole.

Besley, Timothy. 1995. "Property rights and investment incentives: Theory and evidence from Ghana." *Journal of Political Economy* 103(5):903–937.

Besley, Timothy & Robin Burgess. 2000. "Land reform, poverty reduction, and growth: Evidence from India." *Quarterly Journal of Economics* 115(2):389–430.

Binswanger, Hans, Klaus Deininger & Gershon Feder. 1993. "Agricultural land relations in the developing world." *American Journal of Agricultural Economics* 75(5):1242–48.

Binswanger, Hans, Klaus Deininger & Gershon Feder. 1995. Power, distortions, revolt and reform in agricultural land relations. In *Handbook of Development Economics*, ed. Jere Behrman & T.N. Srinivasan. Amsterdam: Elsevier pp. 2659–2772.

Blundell, Richard & Stephen Bond. 2000. "GMM Estimation with persistent panel data: an application to production functions." *Econometric Reviews* 19(3):321–340.

Bommiere, Antoine & Sylvie Lambert. 2004. "Human capital investments and family composition." *Applied Economics Letters* 11(3):193–196.

Brandt, Loren, Jikun Huang, Guo Li & Scott Rozelle. 2002. "Land rights in rural China: Facts, fictions and issues." *The China Journal* pp. 67–97.

Bubb, Ryan. 2011. "States, law and property rights in West Africa.".

Butcher, Kristin F. & Anne Case. 1994. "The effect of sibling sex composition on women's education and earnings." *The Quarterly Journal of Economics* 109(3):531–563.

Cameron, A. Colin, Jonah Gelbach & Douglas L. Miller. 2008. "Bootstrap-based improvements for inference with clustered errors." *The Review of Economics and Statistics* 90(3):414–427.

Chinn, Dennis. 1979. "Team cohesion and collective-labor supply in Chinese agriculture." *Journal of Comparative Economics* 3:375–94.

Coale, Ansley & Judith Banister. 1994. "Five decades of missing females in China." *Demography* 31(3).

Commission, India Government Planning. 1959. "Reports of the Committees of the Panel on Land Reforms.".

Commission, India Government Planning. 1966. "Implementation of Land Reforms: A Review by the Land Reforms Implementation Committee of the National Development Council.".

Conning, Jonathan & James Robinson. 2007. "Property rights and the political organization of agriculture." *Journal of Development Economics* 82:416–447.

Conti, Gabriella, James J. Heckman, Junjian Yi & Junsen Zhang. 2010. "Early health shocks, parental responses and child outcomes.".

Cook, Seth, Li Fengrui & Wei Huilan. 2000. "Rainwater harvesting agriculture in Gansu Province, People's Republic of China." *Journal of Soil and Water Conservation* 55(2):112–114.

Das, Sakmar. 2000. A critical evaluation of land reforms in India (1950-1995). In *Land reform in India: an unfinished agenda.* New Delhi: Sage.

de Janvry, Alain, Marco Gonzalez-Navarro & Elisabeth Sadoulet. 2011. "Can a populist political party bear the risk of granting complete property rights? Electoral outcomes of Mexicos second land reform.".

de la Rupelle, Maelys, Deng Quheng, Qi Shi & Thomas Vendryes. 2009. "Land rights insecurity and temporary migration in rural China.".

De Soto, Hernando. 2000. *The mystery of capital: Why capitalism triumphs in the west and fails everywhere else.* Basic Books.

Deininger, Klaus & Hans Binswanger. 1999. "The evolution of the World Bank's land policy: Principles, experience and future challenges." *The World Bank Research Observer* 14(2):247–76.

Duflo, Esther, Raghabendra Chattopadhyay, Rohini Pande, Lori Beaman & Petia Topalova. 2009. "Powerful Women: Does Exposure Reduce Bias?" *Quarterly Journal of Economics* 124:1497–1540.

Eashvaraiah, P. 1985. *Political dimensions of land reform in India.* New Delhi: Ashish Publishing House.

Feder, Gershon, Lawrence Lau, Justin Lin & Xiaopeng Luo. 1992. "The determinants of farm investment and residential construction in post-reform China." *Economic Development and Cultural Change* 41(1):1–26.

Field, Erica. 2005. "Property rights and investment incentives in urban slums." *Journal of the European Economic Association Papers and Proceedings* 3(2-3):279–290.

Frolking, Steve, Jianjun Qiu, Stephen Boles, Xiangming Xiao, Jiyuan Liu, Yahui Zhuang, Changsheng Li & Xiaoguang Qin. 2004. "Combining remote sensing and ground census data to develop new maps of the distribution of rice agriculture in China." *Plant and Soil* 16(4):89–99.

Galiani, Sebastian & Ernesto Schargrodsky. 2010. "Property rights for the poor: Effects of land titling." *Journal of Public Economics* 94(9-10):700–729.

Garg, Ashish & Jonathan Morduch. 1998. "Sibling rivalry and the gender gap: evidence from child health outcomes in Ghana." *Journal of Population Economics* 11(4):1432–1475.

Ghatak, Maitreesh & Dilip Mookherjee. 2011. "Land Acquisition for Industrialization and Compensation for Displaced Farmers.".

Ghatak, Maitreesh & Sanchari Roy. 2007. "Land reform and agricultural productivity in south India: a review of the evidence." *Oxford Review of Economic Policy* 23(2):251–269.

Goldstein, Markus & Christopher Udry. 2008. "The profits of power: Land rights and agricultural investment in Ghana." *Journal of Political Economy* 116(6):981–1022.

Government of India, Manager of Publications. 1955. "Report on the States Reorganization Commission.".

Grantham-McGregor, Sally & Cornelius Ani. 2001. "A review of studies on the effect of iron deficiency on cognitive development in children." *Journal of Nutrition* 131:649S–668.

Grantham-McGregor, Sally, Yin Bun Cheung, Santiago Cueto, Paul Glewwe, Linda Richter & Barbara Strupp. 2007. "Development potential for children in the first five years in developing countries." *The Lancet* 369(9555):60–70.

Griliches, Zvi. 1979. "Sibling models and data in economics: beginnings of a survey." *Journal of Political Economy* 87(5):S37–S64.

Hazarika, Gautam. 2000. "Gender differences in children's nutrition and access to health care." *Journal of Development Studies* 37(1):73–92.

Herring, Ronald. 1970. Explaining anomalies in agrarian reform: lessons in South India. In *Agrarian Reform and Grassroots Development: Ten Case Studies*, ed. Roy Prosterman, Mary Temple & Mary Handstad. Boulder: Lynne Riener pp. 49–74.

Herring, Ronald. 1991. "From structural conflict to agrarian stalemate: lessons from South India." *Journal of Asian and African Studies* 26:169–188.

Hornbeck, Richard. 2010. "Barbed wire: property rights and agricultural development." *Quarterly Journal of Economics* 125(2):767–810.

Jacoby, Hanan, Guo Li & Scott Rozelle. 2002. "Hazards of expropriation: Tenure insecurity and investment in rural China." *The American Economic Review* 92(5):1420–1447.

Keliang, Zhu, Roy Prosterman, Ye Jianping, Li Ping, Jeffrey Reidinger & Ouyang Yiwen. 2007. "The rural land question in China: Analysis and recommendations based on a seventeen-province survey." *International Law and Politics* 38:761–839.

Kumar, Dharma. 1962. "Caste and landlessness in South India." *Comparative Studies in Society and History* 4(3).

Kung, James. 2000. "Common property rights and land reallocations in rural China: Evidence from a village survey." *World Development* 28(4):701–719.

Levinsohn, James & Amil Petrin. 2003. "Estimating production functions using inputs to control for unobservables." *The Review of Economic Studies* 70(2):317–341.

Lin, Justin Yifu. 1992. "Rural reforms and agricultural growth in China." *American Economic Review* 82(1):34–51.

Martorell, Reynaldo. 1999. "The nature of child malnutrition and its long-term implications." *Food and Nutrition Bulletin* 20:288–292.

Matsuyama, Kiminori. 1992. "Agricultural productivity, comparative advantage and economic growth." *Journal of Economic Theory* 58:317–334.

Mearns, Robin. 1999. *Access to land in rural India.* World Bank.

Michalopoulos, Stelios & Elias Papaioannou. 2011. "Divide and rule or the rule of the divided? Evidence from Africa.".

Mookherjee, Dilip & Pranab Bardhan. 2010. "Determinants of redistributive politics: An empirical analysis of land reforms in West Bengal, India." *American Economic Review* 100(4):1572–1600.

Morduch, Jonathan. 2000. "Sibling rivalry in Africa." *American Economic Review* 90(2):405–409.

National Bureau of Statistics of the PRC. 2006. *China Statistical Yearbook 2006.* China Statistics Press.

North, Douglass & Robert P. Thomas. 1973. *The rise of the Western world: A new economic history.* Cambridge University Press.

Olley, G. Steven & Ariel Pakes. 1996. "The dynamics of productivity in the telecommunications equipment industry." *Econometrica* 64(6):1263–1297.

Ono, Hiroshi. 2004. "Are sons and daughters substitutable? Allocation of family resources in contemporary Japan." *Journal of the Japanese and International Economies* 18:143–160.

Pande, Rohini & Christopher Udry. 2006. Institutions and development: A view from below. In *Advances in economics and econometrics: theory and applications*, ed. Whitney K. Newey & Torsten Persson. Cambridge: Cambridge University Press.

Pani, Narendar. 1983. *Reforms to preempt change: Land legislation in Karnataka.* New Delhi: Concept Publishing Company.

Parish, William L. & Robert Willis. 1993. "Daughters, education and family budgets: Taiwan Experiences." *The Jouanl of Human Resources* 28(24):863–898.

Pollitt, Ernesto, Kathleen S. Gorman, Patrice L. Engle, Juan A. Rivera & Reynaldo Martorell. 1999. "Nutrition in Early Life and the Fulfillment of Intellectual Potential." *Journal of Nutrition* 125:1111S–1118.

Qian, Zhenchao. 1997. "Progression to second birth in China: a study of four rural counties." *Population Studies* 51(2):221–229.

Radhakrishnan, P. 1990. "Land Reforms: Rhetoric and Reality." *Economic and Political Weekly* .

Rajan, M.A.S. 1986. *Land reforms in Karnataka: An account by a participant observer.* New Delhi: Hindustan Publishing Corporation.

Ray, Debraj. 1998. *Development Economics.* Princeton: Princeton University Press.

Rosenzweig, Mark & Kenneth Wolpin. 1988. "Heterogeneity, intrafamily distribution and child health." *Journal of Human Resources* 23(4):427–461.

Rosenzweig, M.R. & H.P. Binswanger. 1993. "Wealth, weather risk and the composition and profitability of agricultural investments." *Economic Journal* 103:56–78.

Rosenzweig, Paul & Junsen Zhang. 2009. "Do population control policies induce more human capital investment? Twins, birth weight and China's 'one-child policy." *The Review of Economic Studies* 76:1149–1174.

Rozelle, Scott & Guo Li. 1998. "Village leaders and land-rights formation in China." *The American Economic Review* pp. 433–438.

Schwarzwalder, Brian, Roy Prosterman, Ye Jianping, Jeffrey Riedinger & Li Ping. 2002. "An update on China's rural land tenure reforms: Analysis and recommendations based on a seventeen-province survey." *Columbia Journal of Asian Law* 16:141–225.

Shah, Ghanshyam. 2004. Introduction: Caste and democratic politics in India. In *Caste and democratic politics in India*, ed. Ghanshyam Shah. New Delhi: Anthem pp. 1–27.

Sharma, K.L. 1984. *The changing rural stratification.* Delhi: Orient Longman.

Srinivas, M.N. 1966. *Social change in modern India.* Bombay: Allied Publishers.

Stiglitz, Joseph. 1988. Economic organization, information and development. In *Handbook of Development Economics*, ed. Hollis B. Chenery & T.N. Srinivasan. Amsterdam: Elsevier.

Tenikue, Michel & Bertrand Verheyden. 2007. "Birth Order, Child Labor and Schooling: Theory and Evidence from Cameroon.".

Trautman, Thomas. 1981. *Dravidian Kinship.* Cambridge: Cambridge University Press.

UNESCAP. 2002. Population and family planning in China by province: Gansu Province. Technical report United Nations Economic and Social Commission for Asia and the Pacific.

Unger, Jonathan. 1985. "The decollectivization of the Chinese countryside: A survey of twenty-eight villages." *Pacific Affairs* 58:585–606.

Unger, Jonathan. 2005. "Family customs and farmland reallocations in contemporary Chinese villages." *Social transformations in Chinese society* 82:113–130.

Walker, Kenneth. 1984. "Chinese agriculture during the period of the readjustment, 1978-83." *The China Quarterly* 100:783–812.

Wan, Guang & Enjiang Cheng. 2001. "Effects of land fragmentation and returns to scale in the Chinese farming sector." *Applied Economics* 33(2):183–194.

Wen, G.J. 1989. "The current land tenure system and its impact on long term performance of farming sector: The case of modern China.".

World Health Organization. 1995. "An evaluation of infant growth: the use and interpretation of anthropometry in infants." *Bulletin of the World Health Organization* 73:163–74.

Zhang, F., J. Shen, L. Li & X. Liu. 2004. "An overview of rhizosphere processes related with plant nutrition in major cropping systems in China." *Global biogeochemical cycles* 260:1–10.